Vol. 46 MAY, 1955 No. 5

The Journal of Educational Psychology

UNIVERSITY OF MICHIGAN
OCT 24 1955
PERIODICAL READING ROOM

CONTENTS

A Drawing Technique for Studying Certain Outcomes of Teacher Education 257
WILLIAM RABINOWITZ AND ROBERT M. W. TRAVERS

Team Work in the College Class 274
HENRY CLAY SMITH

The Relationship of Style Difficulty to Immediate Retention and to Acceptability of Technical Material 287
GEORGE R. KLARE, JAMES E. MABRY, AND LEVARL M. GUSTAFSON

Prediction of the Academic Success of College Freshmen 296
ROBERT A. JACKSON

A Study of the Effects of Teacher Knowledge of Pupil Characteristics on Pupil Achievement and Attitudes towards Classwork 302
KENNETH B. HOYT

Book Reviews 311

Published monthly except June to September
$7.00 a year in U. S. and Pan Amer.; Canada, $7.20; other countries, $7.40.
Single issues $1.20

WARWICK & YORK, INC.
BALTIMORE 2, MD.

Entered as Second Class Matter Nov. 15, 1921, at the Post Office at Baltimore, Md. under the Act of March 3, 1879.

UNIVERSITY OF MICHIGAN LIBRARIES

THE JOURNAL OF Educational Psychology

Established 1910

BOARD OF EDITORS*

WM. CLARK TROW, *Chairman*
University of Michigan

STEPHEN M. COREY
Teachers College
Columbia University

J. B. STROUD
State University of Iowa

*Appointed by the American Psychological Association.

EDITORIAL CONSULTANTS

PAUL BLOMMERS
CHESTER W. HARRIS
HAROLD E. JONES
H. H. REMMERS
PERCIVAL M. SYMONDS
MILES A. TINKER
ALEXANDER G. WESMAN
PAUL A. WITTY

WARWICK & YORK, INC.

Publishers

regretfully announce the death of its beloved President

H. E. Buchholz

on July 25, 1955, after a short illness. The publication of the Journal of Educational Psychology, established by Mr. Buchholz in 1910, and of Educational Administration & Supervision, also established by him in 1915, will be continued without interruption by the Publishers, in coöperation with the Editors.

THE JOURNAL OF

EDUCATIONAL PSYCHOLOGY

Volume 46 May, 1955 Number 5

A DRAWING TECHNIQUE FOR STUDYING CERTAIN OUTCOMES OF TEACHER EDUCATION[1]

WILLIAM RABINOWITZ

Bank Street College of Education, New York City

and

ROBERT M. W. TRAVERS

Personnel Research Laboratory, Air Force Personnel and Training Research Center, San Antonio, Texas

This article deals with a drawing technique and its success in differentiating students exposed to different teacher education experiences. Evidence of the sensitivity of the technique to changes that take place in the course of teacher education will also be presented.

Students at two teacher training institutions at different stages in their professional preparation were given a blank sheet of paper at the top of which were the following directions: "In the space below, draw a picture of a teacher with a class. Draw as complete a picture as you can. Avoid the use of stick figures. Don't worry about your artistic ability or lack of it; just draw as well as you can."

This article is based upon an analysis of the drawings thus obtained.

---

[1] The research described in the article was undertaken and completed while the authors were members of the Staff of the Office of Research and Evaluation, Division of Teacher Education, Board of Higher Education, New York City.

UNIVERSITY OF MICHIGAN LIBRARIES

THE MEANING OF DRAWINGS

There is a growing literature on interpretations of drawings made by clinicians to delineate psychodynamic trends in the cases they study. Drawings have also been used to estimate intelligence in children, especially those with language difficulties. The relevance of these methods of analyzing drawings is marginal in the present investigation. Though drawings have been used as a tool in individual diagnosis our purposes are somewhat different.

If a person is asked to draw a teacher in a class, or to draw a policeman directing traffic, or to draw a foreman supervising a road construction gang, he does not produce a drawing out of a vacuum. The drawings inevitably display some of the ideas he has acquired concerning what these individuals do in these particular situations. This knowledge may be objectively correct or it may be distorted, but that is of no immediate consequence. In a picture of this kind, individuals project their impressions concerning how they think a person might act in the situation specified. Without such projection there could be no drawing. We are not saying that the person who produces a drawing does not exercise some conscious selection over the ideas he presents. In most cases, there undoubtedly is some selection of ideas. However, we have found that the individuals who perform this simple test do not usually appreciate the fact that they reveal their own highly personal ideas about the persons and situations presented.

An aspect of projection which is particularly striking is the fact that in his drawing a student often portrays a teacher who actually looks like himself. A tall teacher is drawn most often by a tall student. Heavy students are likely to draw heavy teachers. The age and sex of the drawer are almost always reflected directly in the drawings. These evidences of projection are most discernible when physical similarities are the object of study. However, projection is not limited to the body image, but extends to a great host of attitudes and conceptions which the student holds.

Many of those with whom the drawing technique was first discussed questioned the procedure since individuals differ greatly in drawing skill. These differences in skill, it was argued, would mean that gifted individuals would have an advantage over those less endowed. Drawing skill would make it possible to represent a teacher in the classroom more completely and attractively. The

partial truth of this objection must be admitted. Drawing skill does apparently give an individual greater control over his production. Yet our experience has been that the control manifests itself most often in ways which are irrelevant from the viewpoint of interpretation. It is unimportant whether an idea is well presented or poorly presented. Drawing interpretation focuses attention upon the content of the ideas, rather than its formal representation. The skilled individual has an advantage in those aspects of drawing performance which are of least importance.

TWO DRAWING EXAMPLES

Figure 1 and Figure 2 show two radically different drawings produced by student teachers from the same institution. Both students were young women, both were planning to teach at the elementary school level, and both had taken the same number of credits in professional courses in education. Yet the differences in the drawings are striking. The one represents an informal program in which different children participate in different activities at the same time, and in which the teacher appears to exercise little control over the movements of the children. The other picture represents a more formally structured classroom situation. It seems that the teacher controls the entire situation, all the children perform the same activity at the same time, and the emphasis is on academic learning.

We do not say at this point that these two student-teachers would behave differently when faced with a class. We say only that the drawings portray classroom situations which are radically different. The extent and nature of the differences in the students who produce them is another matter.

Certain additional differences may be noted in these two drawings. In the one drawing, the teacher is the center of the entire picture as well as being the center of the entire learning procedure. In the other one, the teacher is drawn more to the side of the picture and is attending to only a few of the children. In the former one, the teacher is drawn in greater detail and with more definiteness than the children. In the other drawing, no special emphasis is given to the teacher who is drawn in about as much detail as the children. In the one drawing, the only piece of equipment is the blackboard with material on it representing a formal, academic approach to reading. In the other, there is a wealth of equipment

UNIVERSITY OF MICHIGAN LIBRARIES

FIGURE 1

and the blackboard is used by the children. The extent to which the teacher is represented as physically close to the children differs in the two pictures. In one there is a gulf between teacher and class, but in the other they are close together.

Certain drawings are extremely bizarre, and even to an observer who has no clinical training they seem 'queer'. Most of them portray some activity, but not an integrated meaningful response pattern in the classroom. The sex of the teacher figure may be impossible to identify. Gross distortions in perspective or the size of various body parts may be present. The drawing is usually permeated by confusion. Drawings of this kind were found in only a very small minority of those examined. In three cases it was possible to obtain informal evaluations of the persons involved given by faculty members over several years. In two cases, the impression of personality disturbance gained from the drawing was directly confirmed by the faculty records. The third case was ambiguous.

In a preliminary examination of the drawings, several students were selected whose drawings before and after student-teaching

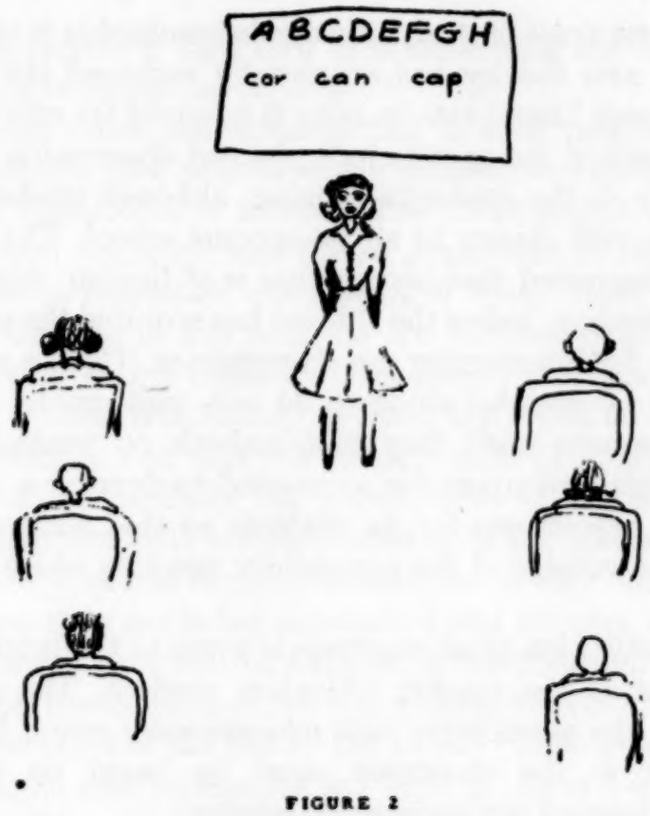

FIGURE 2

showed great changes. Several of these students were interviewed in order to learn if the drawing changes were related to the student-teaching experience. In almost all cases they were.

A QUANTITATIVE APPLICATION OF THE TECHNIQUE

The first part of this paper described a technique for conducting enquiries into the concepts of teaching possessed by an individual. In order to explore further the usefulness of this technique and to investigate the possibility of deriving certain measures from it, the drawings of students of education in two different teacher-education programs were studied. These two programs in Institution A and Institution B differ in many basic respects as is evident from the following descriptions.

DESCRIPTION OF TWO INSTITUTIONS

Institution A.—The program of teacher education in this institution is generally described as being conservative in character. The emphasis on the liberal arts rather than on professional education is reflected in the fact that it does not offer a major in education. The student's major field of specialization must be one of

UNIVERSITY OF MICHIGAN LIBRARIES

the liberal arts areas of study. Academic scholarship is emphasized. The liberal arts faculty has continually endorsed the viewpoint that a thorough liberal arts training is essential for effective teaching. The program does not include planned observation in selected schools early in the student's training, although students are encouraged to visit classes in an on-campus school. The opinion is commonly expressed that observation is of limited value, or even may be misleading, unless the student has acquired the professional background for interpreting the observations. Thus, a majority of the teacher educational students do not visit public schools as student observers until they first embark on student-teaching. However, this institution has attempted to develop a program of community experiences for its students so that they will have a first-hand knowledge of the community agencies which serve children.

In this institution great emphasis is given to the intellectualistic development of the teacher education student. The program is based upon the assumption that educationally sound behavior of the teacher in the classroom must be based on sound and thoroughly learned professional knowledge.

In this program relatively little is done or can be done to weed out those whose personality seems unsuitable for teaching. Although the speech department does all that can be done to eliminate those who display undesirable behavior during a speech test, the observation of the student in a formal academic setting does not permit the careful and detailed study of individual peculiarities.

Institution B.—In many respects, the teacher education program in Institution B is based on different assumptions from that in Institution A. Professional training is given emphasis in the total program of the teacher education student. It is emphasized that the teacher education student is being trained primarily as a teacher although he is expected to be competent in his subject matter field.

A major assumption underlying the program is that knowledge alone does not guarantee that the teacher will manifest a desirable pattern of behavior in the classroom. Thus, considerable effort is made to provide the student with opportunities to observe in selected schools from the earliest weeks of his professional training. Carefully planned field and laboratory experiences form an integral part of the professional training of the student, and student-

teaching represents a final step in familiarizing the student with the problems of educating children.

The schools which the students visit in their programs are not just any schools, but schools which present an educational outlook which Institution B accepts. The development of this program of school visitation and the establishment of relations with suitable schools represents many years of work. Thus, the teacher education students in this program are confronted with a particular pattern of teacher behavior throughout their education.

There is a close relationship between the students and faculty in this institution and this relationship makes it possible for the faculty to study carefully the personalities of the students. On this basis more students are eliminated from the teacher education program. The students themselves seem to be very much aware of the fact that they are being scrutinized and develop anxieties related to whether they will be retained in the program.

These brief descriptions of the two programs should not be taken as an endorsement of the one or the other. Undoubtedly each has its own merits and represents the fruit of serious thought and study of faculties which differ in their values and in their theories concerning how their values can be best achieved. We are not concerned here with problems of value systems but with the relationship of training conditions to concepts of teaching.

PROCEDURE

The basic data in the present study consist of drawings of 'a teacher in a class' made by students in these two institutions who were at various stages of training in the teacher education program. It might have been desirable to follow the same group of students through for a period of two or three years and to have studied changes associated with teacher-training, but the time was not available for undertaking a longitudinal study of this character. Hence, it was necessary to obtain observations on different groups at a number of different levels of training. In order to simplify the study, it was decided to limit it to those planning to become elementary school teachers.

There is no doubt that it would have been desirable to match groups for such variables as age, sex, intelligence, and type of curriculum experienced in high school. In practice, it was, however, not possible to do this. It seems probable that any differences be-

UNIVERSITY OF MICHIGAN LIBRARIES

tween the two groups on these variables have not greatly influenced the data. The following groups were included in this study:

Institution A	*Institution B*
17 students in Education I (mainly upper sophomores)	17 students in Education I (mainly lower sophomores)
24 students before student-teaching	20 students in Education III (mainly lower juniors)
24 students (the same ones as above) at the end of student-teaching	20 students just after student-teaching

The actual number of students whose drawings were chosen at each level is less than the number who supplied drawings. These students were chosen through random selection. It would have been desirable to have employed larger numbers, but the work of quantification at this stage proved too great. The limited number of cases does not, however, seriously lessen the impact of the results obtained.

QUANTIFYING SOME OF THE CHARACTERISTICS OF DRAWINGS

The drawings are complex representations of complex phenomena. A given pattern of teacher behavior may be represented in drawings in an endless number of different ways. Yet, within this diversity of representation, certain common ideas on specific phases of the teacher's behavior in the classroom are nearly always experienced. For example, the teacher may be separated as the central and major element in the entire drawing or may be so hidden among other objects that he is hard to find. Most drawings will fall somewhere between these two extremes and the drawings can be rated for their position along this scale.

Our initial efforts were devoted to a study of the kinds of characteristics with respect to which consistent ratings could be obtained. This involved the preparation of rating sheets related to characteristics of drawings in which we were interested and the use of these rating sheets by several raters who were graduate students in psychology. The raters were later brought together in conference in order to discover the reasons for discrepancies between ratings. The rating sheet was then revised in order to eliminate those aspects of the rating procedure on which reasonable consistency among raters could not be achieved. The revised rating sheet was again subjected to this procedure which was coninued until a rating sheet was developed which seemed adequate

for the present purposes. It may be noted that the raters did not know at any time which group they were rating.

The average ratings for each group are presented as points on a graph in order to show trends during teacher education. The results for the three raters have been plotted separately except for the case of the final graph. Tests of significance were applied (1) in order to evaluate the differences between groups at the beginning and end of training and (2) for the purpose of comparing the students in the two institutions at the end of training.

HOW DOES TEACHER-TRAINING AFFECT THE EXTENT TO WHICH THE DRAWINGS SHOW ACTIVITY OUTSIDE OF THE CONTROL OF THE TEACHER AND ACTIVITY SHOWING PARTIAL OR TOTAL PUPIL CONTROL?

Major differences might be expected between Institution A and Institution B in the extent to which the concepts of teaching of the students graduating from these programs included pupil activity outside of the direct control of the teacher. The drawings themselves demonstrated striking differences in this respect as is evidenced from the graphs shown in Figure 3 and Figure 4.

The graphs present data based upon two somewhat distinct ratings. The first, summarized in Figure 3, involves the percentage of drawings in which any activity outside the immediate control of the teacher is portrayed. For example, if a drawing showed a picture of a class in which a formal lesson was being conducted, but with a few pupils at the side or back of the room engaged in some other activity then that drawing would be classified as showing some activity outside of the teacher's immediate control. The second rating, summarized in Figure 4, involves more than this, and indicates the extent to which the drawings portray situations in which there is partial or total pupil control.

The percentage of drawings for the various groups which shows activity outside of the immediate control of the teacher is shown in Figure 3. Changes associated with teacher training are somewhat greater in Institution B than in Institution A. This is to be expected since the concept of teaching which Institution B attempts to develop emphasizes the importance of allowing pupils to engage in activities without direct and constant teacher supervision.

It seems desirable to apply a test of significance in order to determine the probability that the differences between the ratings

UNIVERSITY OF MICHIGAN LIBRARIES

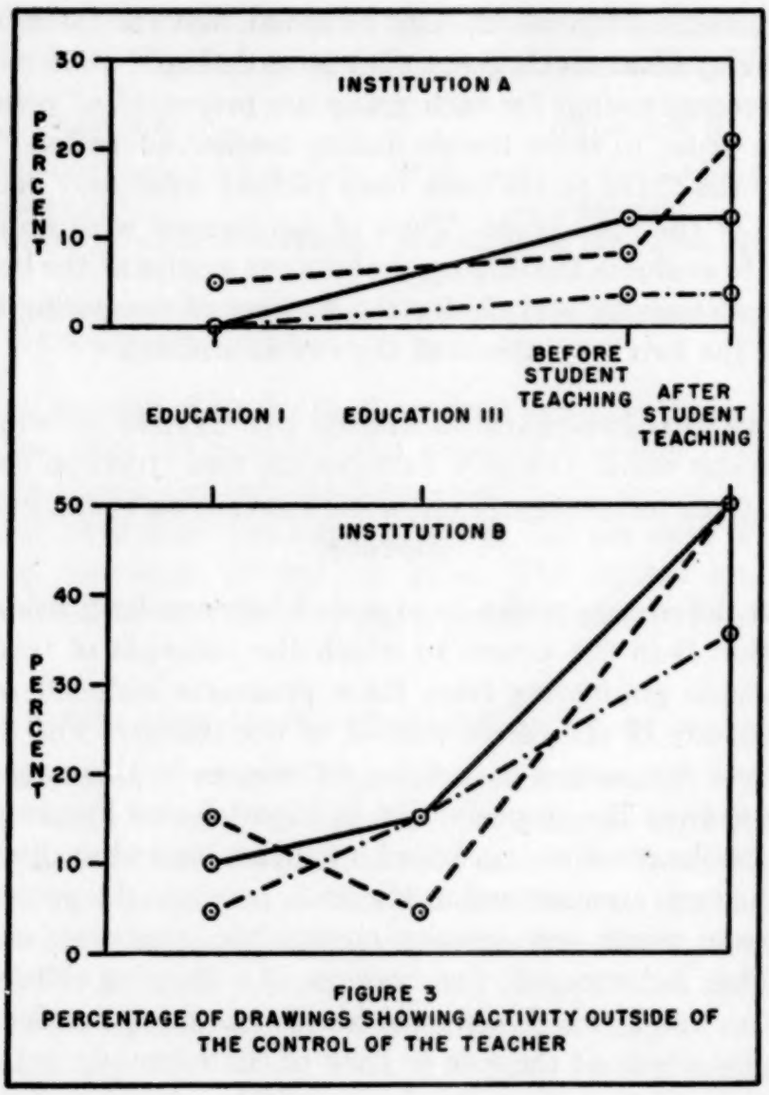

FIGURE 3
PERCENTAGE OF DRAWINGS SHOWING ACTIVITY OUTSIDE OF THE CONTROL OF THE TEACHER

at the beginning and end of training at Institution B are only chance fluctuations. This is done by first drawing up the following four-fold table:

	Number of Drawings Showing Activity Outside of the Control of the Teacher	Number of Drawings Showing No Activity Outside of the Control of the Teacher
Students in Education I	2	15
Students after student-teaching	9	11

It is possible then to apply a chi-square test to determine whether the differences in the way in which the two sets of drawings are

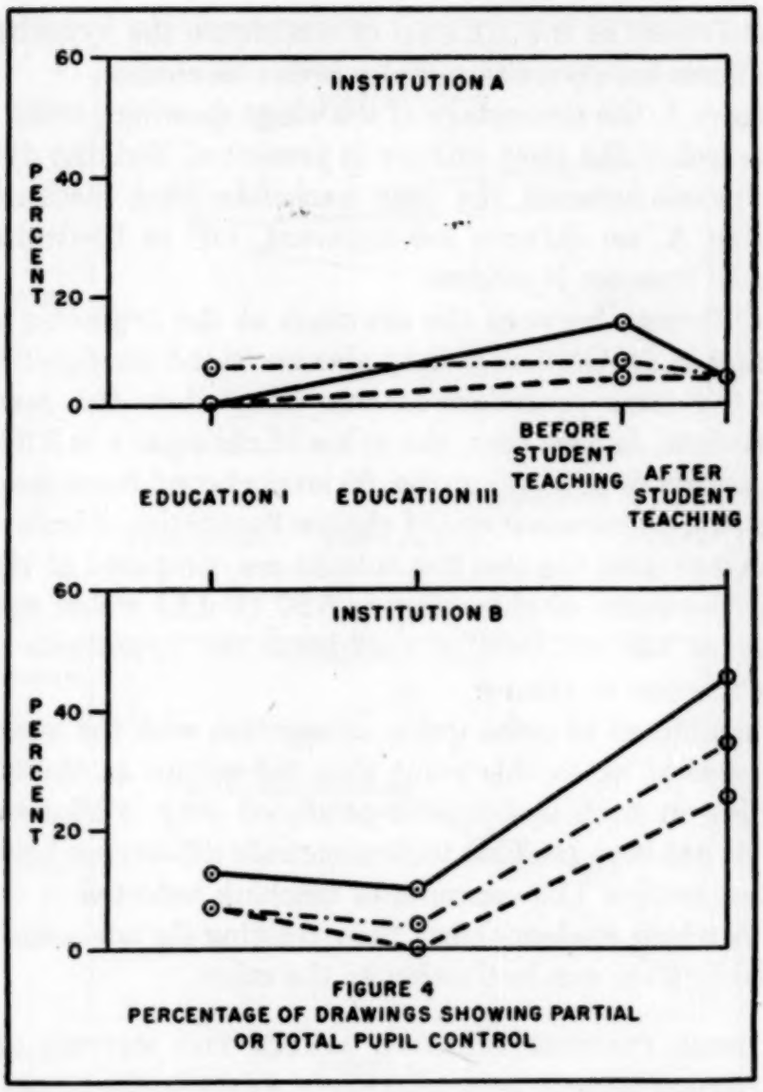

FIGURE 4
PERCENTAGE OF DRAWINGS SHOWING PARTIAL OR TOTAL PUPIL CONTROL

divided should be considered as chance fluctuations.[2] The computed value of chi-square is 4.86 (1 d.f.) which permits us to reject the hypothesis that the differences are a result of chance at the .05 level of confidence.

A similar four-fold table may be drawn up for comparing the drawings made after student-teaching at the two institutions. This table shows a greater difference in the two groups in the fraction of drawings showing activity outside of the control of the teacher. The value of chi-square was found to be 6.94 (1 d.f.) which per-

[2] This four-fold table, and those discussed later, are based on the mean frequencies of the three raters. Yates' correction for continuity was applied in computing all chi-squares.

mits us to reject at the .01 level of confidence the hypothesis that the difference between the samples is due to chance.

In Figure 4, the percentage of drawings showing partial or total pupil control of the class activity is presented. Striking differences will be noted between the two teacher-training institutions. In Institution A, no changes are apparent, but in Institution B, a substantial increase is evident.

The difference between the drawings at the beginning and end of training in Institution B may also be tested for significance by exactly the same procedure as was adopted in the case of the previous data. In this case, the value of chi-square is 3.95 (1 d.f.) which permits us to reject at the .05 level of confidence the hypothesis that the difference is one of chance fluctuation. Similarly when the drawings from the two institutions are compared at the end of training the value of chi-square is 6.97 (1 d.f.) which permits us to reject at the .01 level of confidence the hypothesis that the difference is due to chance.

It is significant to point out in connection with the quantitative data presented up to this point that the groups at the beginning of training in both institutions produced very similar drawings, and it has not been possible to demonstrate differences between the beginning groups. The concepts of teaching reflected in the drawings with which students start their training do not seem to differ appreciably from one institution to the other.

HOW DOES TEACHER TRAINING AFFECT THE NATURE OF THE PUPIL-TEACHER RELATIONSHIPS MANIFESTED IN THE DRAWINGS?

The drawings were rated for two major aspects of pupil-teacher relationships. An attempt was made to provide ratings of the extent to which the pictures portrayed tension in the relationship between pupils and teachers. In addition, the raters were asked to identify those drawings in which teacher and pupils were working together on some project.

The average rating for tension in teacher-pupil relations in the two groups is shown in Figure 5. The contrast between these two graphs is striking. At Institution A, there is practically no change in the amount of tension manifested. In Institution B, there is a striking decline in the amount of tension portrayed. The latter change is probably a result of the fact that pupils in the program

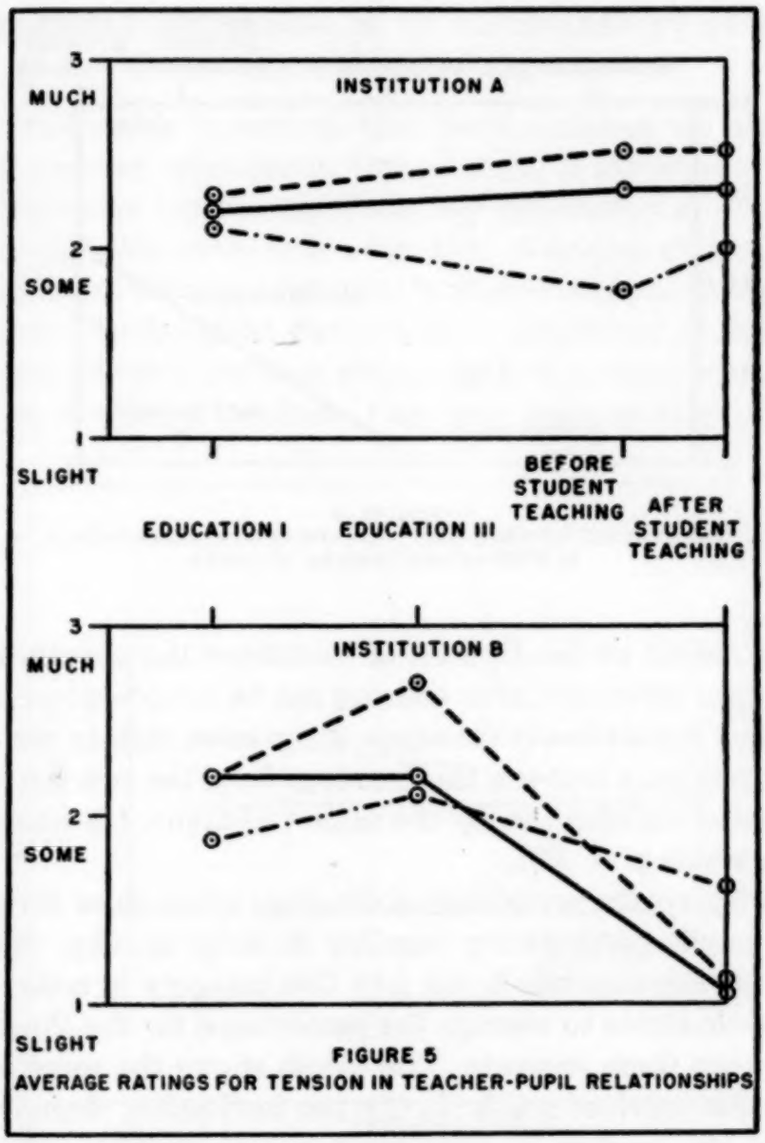

FIGURE 5
AVERAGE RATINGS FOR TENSION IN TEACHER-PUPIL RELATIONSHIPS

have ample and early opportunity to become adjusted to a classroom situation and, hence, to learn adequate responses for handling the situations that arise. At Institution A this process of adjustment can only be begun in student-teaching, and the tensions involved must be resolved in later on-the-job experiences.

In order to test the significance of the difference between the ratings of the drawings at the beginning and end of training at Institution B three ratings for each drawing were averaged. A t test was then applied in order to determine the significance of the difference between the mean ratings for the two groups of drawings. The value of t was found to be 3.546 (d.f. = 35) which per-

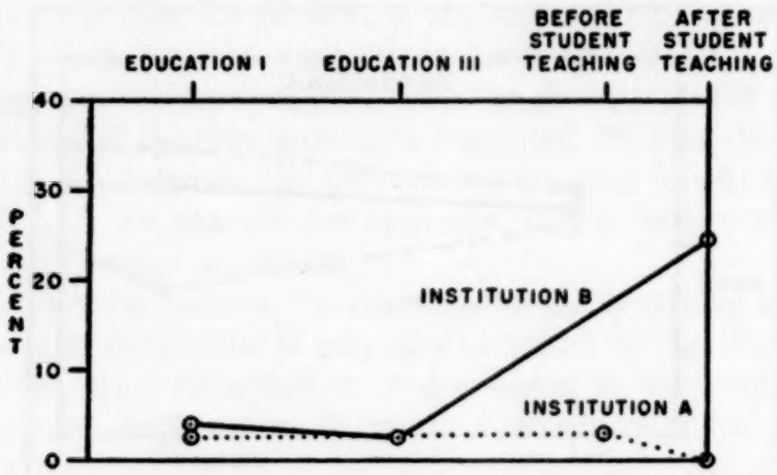

FIGURE 6
PERCENTAGE OF DRAWINGS SHOWING TEACHER AND PUPILS PARTICIPATING IN SOME ACTIVITY

mits us to reject at the .01 level of confidence the hypothesis that the drawings before and after training can be considered as samples of the same population of drawings. Even more certain can one be of the significance between the drawings from the two institutions at the end of training for, by the same technique, t is found to be 4.281 for which $p = .001$.

Figure 6 gives the percentage of drawings which show the teacher and the pupils participating together in some activity. Since the number of drawings which fall into this category is rather small, it seemed desirable to average the percentages for the three raters than to keep them separate. This graph shows the same striking trend as the previous graph. In the one institution, there is an increase in the amount of pupil-teacher participation in common activities; in the other, there is little change and even possibly a decline.

The significance of the difference between the ratings of the drawings made before and after teacher-training at Institution B was determined by the chi-square test in the manner previously described. The value of chi-square for this comparison is 3.82 (1 d.f.) which is significant at the .05 level of confidence. However, one can speak with greater confidence about the difference between the institutions at the end of training with respect to the characteristic rated. In this latter case the value of chi-square is 7.38 (1 d.f.), which is significant at the .01 level of confidence.

VARIABLES HYPOTHESIZED TO BE CONTROLLED BY PROCESSES OTHER THAN TRAINING EXPERIENCES

It is reasonable to assume that some variables are affected by teacher-training experiences. One of them is the extent to which the prospective teacher manifests any psychological disturbance. Presumably, this variable is a function of deep-seated personality characteristics which are not greatly altered in the course of teacher education. Psychological disturbance is manifested in drawings in numerous different features which must be interpreted largely on the basis of clinical intuition. One may presume that the college

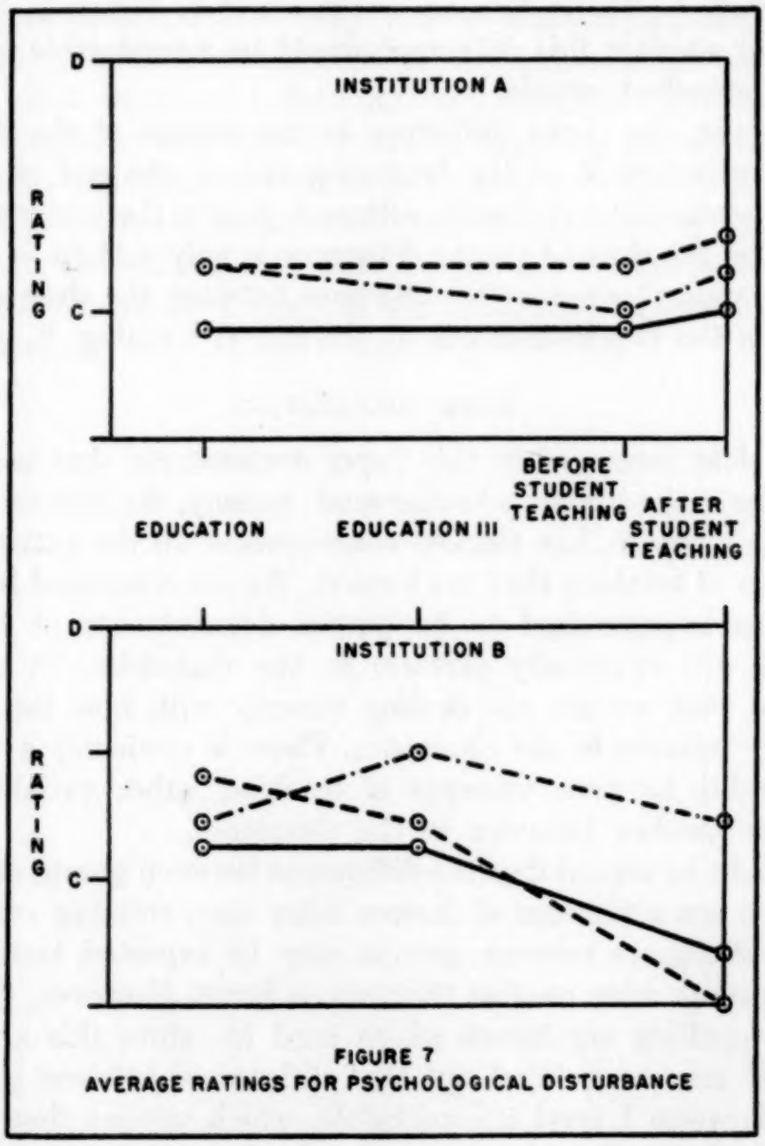

FIGURE 7
AVERAGE RATINGS FOR PSYCHOLOGICAL DISTURBANCE

can exercise some control over the psychological make-up of its students through a program of continuous selection.

It has been pointed out that there are greater opportunities in Institution B than in Institution A for eliminating from the program those who have psychological disturbances which might interfere with classroom teaching. We would thus expect that advanced students of education in Institution B would show a lower incidence of psychological disturbance than corresponding students in Institution A.

In Institution B, advanced students show a slightly lower incidence of psychological disturbance than advanced students in Institution A. These data are summarized in Figure 7, but it is doubtful whether this difference would be reproducible in a new and independent sample.

However, the slight difference in the ratings of the drawings from Institution B at the beginning and at the end of training has no statistical significance although it is in the expected direction. The calculated t for the difference is only 1.41 ($p = 0.2$). Of even less significance is the difference between the drawings produced in the two institutions at the end of training.

SOME CONCLUSIONS

The data presented in this paper demonstrate that one of the variables in the teacher's background; namely, the type of teacher-training program, has marked consequences on the nature of the concepts of teaching that are formed. We are concerned here with variables hypothesized to be partial determinants of how the teacher will eventually perform in the classroom. It must be stressed that we are not dealing directly with how the teacher actually behaves in the classroom. There is obviously a complex relationship between concepts of teaching, other variables, and ultimate teacher behavior in the classroom.

It might be argued that the differences between groups shown by the data are a function of factors other than training conditions. Some differences between groups may be expected because different groups were used at the various levels. However, there are two compelling arguments which tend to refute this argument. First, it may be pointed out that differences between groups at the Education I level are negligible, which implies that prior to training the students in the two institutions are comparable with

respect to the variables studied. Second, it must be noted that differences between the education students which appear as training proceeds are entirely in accordance with expectation. Consistency between trends which appear in the groups and the nature of the teacher education program is so great that it is extremely unlikely that these apparent trends could be a product of irrelevant variables resulting from the sampling procedures.

The data also illustrate how some aspects of concepts of teaching may be studied and how, through such study, the results of the teacher education process can be better understood. Drawings represent an admittedly imperfect instrument and will require much improvement before they can be used as anything more than crude research instruments. It seems important to stress, however, that in spite of their crudity, drawings have reflected certain changes in concepts of teaching following teacher-training in different programs. Drawing as a research technique holds great promise and warrants the attention and study necessary to realize its richness. In addition, new instruments are needed in order to supplement the kinds of material which have so far been successfully elicited.

TEAM WORK IN THE COLLEGE CLASS

HENRY CLAY SMITH

Department of Psychology, Michigan State College

The purpose of the present experiment was to test a set of related hypotheses regarding teamwork. Previous studies by the author and his colleagues have explored nondirective and democratic leadership in the college class (*3, 9*). The former was based upon the conceptions of Rogers (*8*); the latter upon the studies of Lewin and his students (*5*). The present study takes its departure from Nelson's analysis of the integrative leader (*7, 11*).

The integrative leader sees his primary purpose as the development of teamwork and perceives his authority as resting upon informal group codes. It was assumed in the present study that the success of such a leader would be determined by his skill in applying the principles of integrative leadership. A set of tentative principles was inferred from a consideration of laboratory and field studies of group behavior (*2, 4, 5, 6, 12*). Specifically it was assumed that the effective integrative leader would: 1) create work for teams; 2) fit the teams together; 3) encourage interaction on the teams; 4) stabilize the teams; 5) reward the teams as well as the individuals; and 6) let the teams decide.

These assumptions concerning the activities of the integrative leader provided the framework from which the set of hypotheses tested in the present study was derived. These hypotheses are numbered below in accordance with the principle to which they are most closely related.

1) Students in a team class are more satisfied and more productive than students in a lecture class.

2A) Students who like to participate are more satisfied and productive in a team class than those who do not like to participate.

2B) Students who are initially better informed about the content of the course are more satisfied and productive in a team class than those who are not.

2C) Students who are low in manifest anxiety are more satisfied and productive in a team class than those who are high in anxiety.

4) The longer students continue as a team, the more productive they become.

5) Students are more satisfied with a grading procedure which rewards group success than with a procedure which rewards only individual success.

6) Students perceive the instructor playing an integrative rôle as warmer, more objective, and more democratic than when he plays the rôle of lecturer.

It was hoped that the simultaneous testing of these hypotheses would clarify the general problem of teamwork as well as its specific applications to the classroom.

METHODS

The general design of the experiment is shown in Table 1. The ninety-six subjects were students in two general psychology classes. Both classes took a ninety-item achievement test and completed a Participation Scale during the first two days of the course. This latter scale was designed to measure student attitudes toward participation in class discussion. On the basis of these test results, the class chosen for the team experiment was subdivided into five groups. These groups eventually named themselves: 'Psychologists Anonymous' (PA), 'Our Gang' (OG), 'Rearguard' (RG), 'Bow Peeps' (BP) and 'Junior Psychologists' (JP). The students were informed that the groups had been equated for initial achievement. However, they were not informed that the groups also varied in their average attitude toward participation from very favorable to very unfavorable.

The forty-eight students in the team class were individually matched on achievement and attitude with one of the sixty students in the lecture group. As a result of this matching, there were individual and group controls within the lecture class. However, no attention was paid to the groups during the lecture class. Students were seated alphabetically and the class procedures were of the informal lecture-discussion type.

TESTS AND SCALES

Achievement tests.—The initial and final tests of achievement consisted of equivalent forms. Each form was composed of ninety multiple choice items: thirty vocabulary items, thirty factual items, and thirty items involving the application of psychological facts and principles. Statistical data on the reliability of initial and final forms as well as for gains scores have been reported previously (*3*).

TABLE 1.—GENERAL EXPERIMENTAL DESIGN

	Team Class (N-48)			Lecture Class (N-48)		
Pretests:	90-item achievement test and participation attitude scale					
Groups by participation attitudes	N	Scores: Achievement	Scores: Participation	N	Scores: Achievement	Scores: Participation
Very favorable (P.A.)	10	36.4	39.5	10	36.3	38.3
Favorable (O.G.)	9	37.3	36.8	9	37.0	36.0
Average (R.G.)	10	36.0	33.6	10	35.9	33.6
Unfavorable (B.P.)	9	35.9	30.1	9	35.7	30.4
Very unfavorable (J.P.)	10	36.0	25.8	10	35.9	26.4
Average	48	36.3	33.2	48	36.2	32.9
Methods:	Organized into five teams with group projects, group decisions, and a group incentive			Traditional lecture-discussion class with individual grades as the incentive		
	Both classes had same tests and quizzes. Both completed Anxiety scale.					
End Tests:	Equivalent form of initial achievement test. Scales measuring attitudes toward the course, the incentive system, and the instructor.					

Quizzes composed of similar items were given at the end of the third, sixth, and ninth week of the course. They consisted of forty, fifty, and forty items, respectively.

Participation scale.—The participation scale consisted of the following questions:

1) Do you like to participate in class discussions?
2) Do you find it easy to speak in class?
3) Do you think that class discussions are valuable?
4) Is it easy for you to give an oral report before a class?
5) Do you volunteer answers to questions raised by the instructor?

6) Do you volunteer ideas to start discussions in class?

7) Do you do better in oral quizzes than in written ones?

8) Do you do better than most students when called on unexpectedly?

9) How often have you actually participated in your classes?

10) Do you raise an objection when you disagree with the instructor or a fellow student?

The students were asked to indicate their agreement or disagreement with each question on a five-point scale. Thus the minimum score on the scale was 10, the maximum, 50. The corrected odd-even reliability of the scale was .86.

Course satisfaction scale.—This scale, administered and scored in the same way as the participation scale, consisted of the following five items:

1) The classes in this course have been extremely unpleasant.

2) I found the classes dull and uninteresting.

3) I am dissatisfied with the way this course was run.

4) I never became so involved in the classes that I forgot about everything else.

5) The classes were rather uninteresting.

Incentive system evaluation.—This scale consisted of these items:

1) The grading system used in this course should be discontinued.

2) I did not like the grading system used in this course.

3) I think the grading system used in this course was unfair.

4) I was quite dissatisfied with the method of determining grades.

5) The grading system is not as good as that used in other courses.

Anxiety scale.—The anxiety scale utilized in this experiment was originally developed by Spence and Taylor (*10*). It consists of fifty items which were judged as being indicative of anxiety with fifteen additional items introduced to conceal the anxiety items.

The instructor rating scale.—This scale was constructed to measure the degree to which the students perceived the instructor as warm, objective, and democratic.

It was based on the factor analysis of the dimensions of parent-child relations reported by Baldwin (*1*). It consisted of sixty items adapted from Baldwin's interview rating scales. The first twenty attempted to measure the warm-cold dimension, e.g., "He pays considerable attention to the interests of the students". The second twenty measured the objectivity-emotionality dimension, e.g., "He

usually maintains a calm, objective attitude toward the student even in the fact of strong tension". The third twenty measured the democratic-authoritarian dimension, e.g., "He attempts to adjust policies of the class to student wishes wherever possible". The three-dimensions were scored separately. The maximum possible score for each was twenty, the minimum, zero.

CLASS PROCEDURES

In the team class, the five groups were seated in different parts of the room. To encourage group interaction initially, the names of students by groups were mimeographed and passed out and students were given 5 x 8 name cards which each student placed before him during the early meetings.

Grading procedures were also designed to develop group cohesion. An individual grade incentive was used, but the class was informed that each member of the team having the highest average scores on the tests during the course would receive an additional five points on his grade; the next group, four; and so on. Also each team was informed that it would have ten grade points for its group which it could distribute as it decided to the members of the group. Approximately half of their meetings were devoted to projects, designed to further encourage group activity: The determination of reaction time by a group experiment, multiple rôle-playing within the groups, the completion of tests and scales by the group rather than by the individuals, the discussion of case studies, etc. The groups were also permitted to select their own leaders, keep their own attendance records, manage their own class discussions and conduct meetings of their group outside the class. Approximately half of the team class meetings were conducted like those in the lecture-discussion class. A special effort was made to equate the content and method of presentation in these meetings. The lecture class was entirely of the informal lecture-discussion type. The examinations in the two classes were always on the same day.

RESULTS

Tables 2 and 3 present the data used in testing hypotheses 1 and 2A. Table 2 presents the average course satisfaction scores for those groups with the most, intermediate, and least favorable attitudes toward class participation. Table 3 presents the average achieve-

TABLE 2[1].—PARTICIPATION ATTITUDES AND MEAN COURSE SATISFACTION
(Total N-96)

Class	Participation Groups			Total
	V.F. (P.A.)	Intermediate (OG, RG, BP)	V.U. (J.P.)	
Team	20.3	16.3	18.4	17.6
Lecture	18.2	17.4	18.3	17.8
Difference	2.1	−1.1	.1	−.2
Difference of Diff	3.2*		1.2	

[1] In this and later tables the following symbols are used:
* 10 per cent level of confidence
† 5 per cent level of confidence
‡ 1 per cent level of confidence

TABLE 3.—PARTICIPATION ATTITUDES AND MEAN ACHIEVEMENT GAINS

Class	Participation Groups			Total
	V.F. (P.A.)	Intermediate (OG, RG, BP)	V.U. (J.P.)	
Team	18.7	27.0	21.1	24.0
Lecture	25.9	21.1	22.6	22.4
Difference	−7.1*	5.9†	−1.5	1.6
Difference of Diff.	13.0‡		7.5	

ment gains in a similar way. The gains were determined by subtracting the score on the initial test from the score on the final.

As indicated in Table 1, the experiment was designed to contrast the effects of a team class and a lecture class. Thus, in these and the following tables, scores in the lecture class are subtracted from those in the team class in an attempt to isolate the influence of the team situation. In Table 2, for example, the average course satisfaction of the forty-eight students in the team class was 17.6, in the lecture class, 17.8. The difference, −.2, indicates that the lecture class was slightly more satisfying than the team class. The comparable figure in Table 3 was 1.6, which indicates the

average student in the team class gained 1.6 more points than the average student in the lecture class.

These results lend no support to the hypothesis that team work is more satisfying. They lend little support to the hypothesis that teamwork is more productive. In the light of additional facts the satisfaction results appear somewhat more favorable. Table 7 indicates that the team class was dissatisfied with its group incentive system. Further, the correlation between course satisfaction and satisfaction with the incentive system was .46 in the team class and .22 in the lecture class. It seems quite probable, therefore, that with an equally satisfying grading procedure, the team class would have been more satisfied than the lecture class. In regard to achievement, Table 6 indicates that the achievement superiority of the team class became greater as the course proceeded. Thus, more confidence can be placed in the achievement gains on the total course as one point in a trend than in the one score alone. In any case, the insignificant differences in average results are related to the contradictory trends within the class as indicated by the detailed analyses which follow.

Participation attitudes.—Tables 2 and 3 also present the data which were used in testing hypothesis 2A regarding attitudes toward participation. The hypothesis 2 is obviously not supported by the data. Although the very favorable group (PA) was more satisfied in the team class, their achievement gains were less. Actually the Psychologists Anonymous were the most satisfied but made the least achievement gains of any of the ten groups in the experiment. The intermediate groups also gave contradictory results: Although they reported less satisfaction, they made significantly higher achievement gains than either their controls or the very favorable group. The team group with the least favorable attitude (JP) was most similar in satisfaction and achievement to its control group.

The results in regard to the very favorable group were consistent with the instructor's observations that the group, while talkative, were uncoöperative and individualistic. These results and observations reinforce the results of a previous study (*9*) which showed participants to be more dominating, independent, and nonconforming.

High and low initial achievement.—Hypothesis 2B implies that students who made high scores on the initial achievement test

TABLE 4.—INITIAL ACHIEVEMENT IN RELATION TO COURSE SATISFACTION AND ACHIEVEMENT

(N-60)

Class	Course Satisfaction		Achievement Gains	
	High Init. Ach.	Low Init. Ach.	High Init. Ach.	Low Init. Ach.
Team (N-30)	17.5	17.3	24.3	24.7
Lecture (N-30)	18.5	17.6	24.1	22.3
Difference	−1.0	−.3	.2	2.4
Diff. of Diff.	−.7		−2.2	

would perform better in the team situation than those who made low scores.

Fifteen high and fifteen low achievers in the team class were used to test this hypothesis. They were selected by picking the three highest and the three lowest achievers from each of the five groups. The lecture controls were the individuals originally matched with each of these thirty. The results, summarized in Table 4, do not support the hypothesis. The negative differences are insignificant, but it can be safely concluded that the low achievers were not less satisfied or less productive than the high achievers.

High and low anxiety.—Hypothesis 2C stated that anxious students would be less satisfied and lower in achievement than the unanxious ones. The thirty subjects employed in this check were picked in a similar way to those in Table 4. The results, shown in Table 5, contradict the hypothesis: the anxious students in the team class were somewhat more satisfied and made somewhat larger achievement gains. None of the differences are significant, but there is no evidence that the anxious students were less satisfied or less productive as a result of being in the team situation.

Duration of the Groups.—Hypothesis 4 was based on the assumption that it would take time for the team groups to become efficient. In addition to the initial and final tests, both team and lecture groups took three quizzes during the quarter. The data from these tests were used to test the hypothesis. Table 6 reports the mean scores for the team and lecture groups on the three quizzes and the final test. The results are consistent with the hypothesis: the team

TABLE 5.—ANXIETY IN RELATION TO SATISFACTION AND ACHIEVEMENT (N-60)

Class	Course Satisfaction		Achievement Gains	
	Low Anxiety	High Anxiety	Low Anxiety	High Anxiety
Team (N-30)	18.7	16.6	27.2	24.7
Lecture (N-30)	19.4	15.5	23.1	19.4
Difference	−.7	1.1	4.1	5.3
Diff. of Diff.	−1.8		−1.2	

TABLE 6.—ACHIEVEMENT IN RELATION TO NUMBER OF MEETINGS (N-96)

	Test Results After:			
	12	24	35	44 Meetings
Team	28.5	34.5	32.6	59.8
Lecture	29.1	34.7	32.5	58.2
Difference	−.6	−.2	.1	1.6

class improved from one test to the next. The differences are small but consistent. If it is assumed that the team class would get a lower score on the first test but improve regularly thereafter, the chances of obtaining four differences consistent with this hypothesis would be seven in a hundred.

The incentive system.—Hypothesis 5 stated that students would be more satisfied with a grading system which involved a group incentive than with one which only rewarded the individual. The average differences for the entire group on the right of Table 7 indicate that this hypothesis must be rejected at less than the one per cent level of confidence: students in the lecture class were significantly more satisfied with the individual incentive system than the students in the team class were with the group incentive.

In Table 7 the groups are arranged according to their success under the incentive system. The group (OG) in the team class having the greatest success and getting the greatest reward under the

TABLE 7.—ATTITUDES TOWARD GROUP AND INDIVIDUAL INCENTIVES

	Most Successful (OG-9)	Intermediate Success (RG, BP, PA-29)	Least Successful (JP-10)	Total (48)
Team	17.7	17.9	15.7	17.4
Lecture	19.8	21.8	22.1	21.4
Difference	−2.1	−3.9‡	−6.4†	−4.0‡

system was least different from its control in satisfaction with the system. The intermediate groups (RG, BP, PA) were next. The (JP's) were least successful and least satisfied. The difference between these latter two groups and their controls were statistically significant. Students apparently liked the group incentive system more when they profited by it.

Perception of the instructor.—On the basis of Baldwin's tentative findings (*1*) it was assumed that the most effective instructor is one who is warm, objective, and democratic. However, the study was not designed to test this hypothesis. The instructor attempted to be equally as warm, objective, and democratic in the two classes. The data in Table 8 indicate that his efforts were partially successful: in comparison with the lecture class, the team class rated the instructor as significantly more democratic but less objective and slightly less warm. It seems clear that the same instructor in a team class will be judged as more democratic than when he is in a lecture class.

The perception of the instructor's warmth was related to course satisfaction. The correlation between satisfaction with the course and the rating of warmth of the instructor was .51 for the two groups combined (N-96). The variability in the ratings of warmth of the

TABLE 8.—INSTRUCTOR RATINGS

(N-96)

Class	Warmth	Objectivity	Democracy
Team	15.4	13.9	13.3
Lecture	16.2	14.9	11.7
Difference	−.8	−1.0*	1.6‡

same instructor in the same class was great, varying in the team class from 9 to 20; in the lecture class, from 6 to 20. Differences in anxiety may have accounted for part of this, but not all, since the correlation between satisfaction with the course and anxiety was only −.24 and the correlation between anxiety and warmth, only −.19.

DISCUSSION OF RESULTS

The implications of these results are stated as revised hypotheses with which they are consistent, but do not prove.

1) *Work organized for either individual or group effort is more effectively done by groups.* The results of the present study only suggest that college classes which apply the principles of integrative leadership are more satisfying and productive. The inconclusive findings are not surprising when one considers that the procedures were unusual for both instructor and students. A more definitive test requires not only more experienced instructors but also more suitable materials for such courses. In the meantime, the results are sufficiently encouraging to warrant additional explorations.

2) *Groups organized by sociometric procedures are more satisfying and productive than those which are not so organized.* This hypothesis is consistent with the negative results of the present study regarding achievement and anxiety. Common beliefs about the kind of students that fit best into a team situation were not supported: those high in participation, high in initial achievement, and low in anxiety do not profit more than their opposite. Favorable attitudes toward class participation, the most socially relevant of the variables, was most significantly related to success. However, this relationship was not linear: those who were most favorable to class participation did not do so well as those with less extreme attitudes. It seems probable that group success was more likely to be related to social than to personal variables. Considering the numerous positive findings utilizing sociometric procedures, this hypothesis is advanced with some confidence.

3) *Group incentives based on work standards are more satisfying and effective than group incentives based on competition between groups.* The most confident finding of the present study was that students in the team class did not like the group incentive system. Dislike was related to failure under the system. It seems evident that a team would generally be dissatisfied with a system which

penalizes it. However, the typical group incentives in industry are based on production criteria, not directly on competition between groups. Testing this hypothesis would require the comparison of performance under achievement norms set in advance versus the type of competitive group incentive used in the present study.

4) *A warm objective and democratic leader is more effective than one without these qualities.* Some of the present evidence is consistent with this hypothesis. The team class in which the students rated the instructor as more democratic, made somewhat higher scores. Variations in the rating of warmth of the same instructor were significantly correlated with satisfaction with the course (.51).

SUMMARY

A 'team' class in general psychology with forty-eight students was divided into five subgroups. The subgroups were equated for initial achievement but were varied systematically in their attitudes toward class participation. Each member of the team class was matched on the basis of initial achievement and participation attitude with a student in a lecture class taught by the same instructor.

There was practically no difference in the over-all satisfaction of the two classes. However, the team class made greater achievement gains than the lecture class, and the trend in four different tests was significantly upward. In comparison with the lecture class, the team class was significantly more dissatisfied with their group incentive system; the more unsuccessful the subgroup in the team class, the more dissatisfied it was.

The groups with moderately favorable attitudes toward participation made significantly greater gains than their controls in the lecture class. They also made significantly greater gains than the group with the very favorable attitudes toward participation. Although the very favorable group reported the greatest satisfaction, they made the smallest achievement gains. Compared to their controls in the lecture class, students with low anxiety and high initial achievement scores in the team class were less satisfied and made smaller achievement gains than those with high anxiety and low initial achievement.

REFERENCES

1) A. L. Baldwin, J. Kalhorn, and F. H. Breese. *Patterns of Parent Behavior*. Psychol. Monogr., 58, No. 3.

2) L. Coch, and J. R. P. French, Jr. "Overcoming resistance to change." *Human Relations*, 1948, 1, No. 4, 512–532.

3) D. M. Johnson, and H. C. Smith. *Democratic Leadership in the College Classroom*. Psychol. Monogr., 1953, 67, No. 11.

4) H. J. Leavitt. "Some effects of certain communication patterns on group performance." *J. Abnorm. & Soc. Psych.*, 1951, 46, 38–50.

5) K. Lewin, R. Lippitt, and R. K. White. "Patterns of aggressive behavior in experimentally created social climates." *J. soc. Psych.*, 1939, 10, 271–301.

6) E. Mayo, and G. F. F. Lombard. *Teamwork and Labor Turnover in the Aircraft Industry of Southern California*. Harvard University Business Research Studies, 1944, No. 32.

7) C. W. Nelson. A *New Approach to Leadership: Its Rationale, Data, Reliability, Validity, and Application*. Chicago: Industrial Relations Center, University of Chicago, 1950.

8) C. R. Rogers. *Client-centered Therapy*. Boston: Houghton Mifflin, 1951.

9) H. C. Smith, and D. S. Dunbar. "The personality and achievement of the classroom participant." *J. ed. Psych.*, 1951, 42, 65–84.

10) K. W. Spence and J. Taylor. "Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning." *J. Exp. Psych.*, 1951, 42, 183–188.

11) J. J. Valenti. "Measuring educational leadership attitudes." *J. Appl. Psych.*, 1952, 36, 36–44.

12) C. R. Walker and R. H. Guest. *The Man on the Assembly Line*. Cambridge: Harvard University Press, 1952.

THE RELATIONSHIP OF STYLE DIFFICULTY TO IMMEDIATE RETENTION AND TO ACCEPTABILITY OF TECHNICAL MATERIAL[1]

GEORGE R. KLARE, JAMES E. MABRY, AND LEVARL M. GUSTAFSON

University of Illinois[2]

This study was one of a series[3] on the relationship of various communication variables to the learning of technical training material. 'Style difficulty', as used here, can be defined operationally in terms of the aspects of writing measured by readability and level of abstraction formulas. There are many formulas that attempt to measure the readability of writing, and many reports of their application to general reading materials (see *5* and *6* for a summary). There are, also, many articles suggesting the use of formulas.

Most published studies on the value and validity of formulas have used such measures as readership. Few have used retention or related measures (*6*, *9*), and the one study done on technical material[4] yielded negative results. This study was an attempt to relate style difficulty to immediate retention test score, to amount read in a given time, and to acceptability of technical training material.

---

[1] This research was supported in part by the United States Air Force under Contract No. AF 33(038)-25726, monitored by the Personnel and Training Research Center. Permission is granted for reproduction, translation, publication, use and disposal in whole and in part by or for the United States Government.

[2] Dr. Klare is now at Ohio University; Dr. Mabry at VA Hospital, Salt Lake City, Utah; and Dr. Gustafson at Oklahoma A & M College.

[3] For a more complete presentation of this and other studies in the series, see (7).

Grateful acknowledgement is made of the contributions of the following persons to this series of studies: (a) Dr. L. H. Lanier and Dr. L. M. Stolurow, University of Illinois; (b) Mr. E. K. Farson, Mr. F. E. Gregory, Mr. J. H. Orr, Mr. T. E. Carter, Mr. C. S. Hershey, and others, of Chanute Air Force Base, Illinois; and (c) Dr. D. Dubois and Lt. M. Grunzke, Sampson Air Force Base, New York.

[4] "The effect of reading ease on student comprehension of training materials." Hq., Sheppard AFB, 3750 Tech. Trg. Wing. (Mimeo.)

METHOD

Experimental Materials.—A 1201-word multilith-printed lesson from an aircraft mechanics training course was used in this study. The first half of the lesson covered the 'induction system' and the second half the 'cooling system' of the engine. In preparing experimental versions, only the style difficulty, or 'structure', of the material was changed. Technical experts, serving as judges, determined that neither 'content' nor technical terms were changed in any experimental versions developed.

Three 'styles', or levels of style difficulty, were used, 'Easy', 'Present' (the standard version used in the course), and 'Hard'. These styles varied in (a) percentage of short, familiar, frequently used words, (b) average sentence length and the related proportions of simple sentences and prepositional phrases, (c) the proportion of abstract to concrete wordings, and (d) the proportion of active to passive constructions. The 'Easy' style was at about seven to eight grade level and the 'Hard' at about sixteenth plus grade level, as measured by the Dale-Chall readability formula (*1*) and the Flesch readability (*2*) and 'level of abstraction' (*3*) formulas. The 'Present' style was at about eleven to twelve grade level as measured by the Dale-Chall and Flesch readability formulas, and thirteenth to fifteenth grade level by the level of abstraction formula.

Nine 'versions' of the lesson were used in the study. Three were called 'unsplit', treating both halves of the lesson in one style; these were E (for Easy), P (for Present), and H (for Hard). Six versions were called 'split', treating the two halves of the lesson in a different style; these were EP (first half Easy, second half Present), PE, EH, HE, PH, and HP. All versions had the same number of words (1206), and the same format.

Every fifth line of each version was numbered in order to help in getting a measure of amount read by the subjects. Each subject made a tally mark for each complete reading during the experimental period, and indicated the line he was reading when asked to stop.

A fifty-item multiple-choice test was constructed to measure immediate retention. The items were selected from a pool of one hundred and twelve items in such a way that each paragraph of the material was allotted a percentage proportional to its size. The base for writing the items was the Present version of the lesson,

so that any bias which might exist would be in the conservative direction (of course, holding content constant in all versions should have tended to minimize bias). The split-half reliability of the test, corrected by the Spearman-Brown formula, was .87.

Acceptability of the styles was determined by the answers to three questions. The first question asked subjects if they noticed a difference in the way the two halves of the lesson were written (that is, in style, not in what was written about, or content). Those who noticed a difference were then asked which half of the material they thought 'easier to read' (question 2) and which half 'more pleasant to read' (question 3). These questions were asked of those subjects who had the unsplit as well as those who had the split versions.

Subjects and Procedure.—Subjects were nine hundred and eighty-nine male airmen in indoctrination training at Sampson Air Force Base (one hundred and nine subjects for PE, and one hundred and ten for all other versions). Various aptitude indices were available for all airmen, based on the Airman Classification Test Battery (these indices, based on a scale of 1 to 9, are conveniently called 'stanines').

Subjects were given twenty minutes to read the lesson material, and during that period indicated the amount read as previously described. Subjects then answered the three questions designed to determine acceptability of the material. After this, forty minutes were allowed for answering the test.

RESULTS

Amount Read (20 Minutes).—Analyses of these data showed that the Hard version resulted in fewer lines read than the Easy or Present versions. The differences, however, were not significant.

The number of words read, on the other hand, differed more. This could occur because, even though all versions had the same number of words, the Easy style occupied fewer lines than the Present, and the Present fewer than the Hard. It was found that the Easy and Present styles resulted in significantly more words read than the Hard, beyond the .01 level. The Easy style resulted in more words read, but not quite significantly more, than the Present. Table 1 presents relevant data for amount read.

Acceptability.—The next analysis involved the three questions asked of all subjects ("did the two halves differ," "which half easier

TABLE 1.—COMPARISONS OF AMOUNT READ IN TWENTY MINUTES FOR THREE LEVELS OF STYLE DIFFICULTY

Version			Means of Reading Measures		
Level	Lines	Words	Times read*	Lines	Words
Easy	126	1206	2.16	272.25	2605.70
Present	134	1206	2.06	276.65	2489.85
Hard	140	1206	1.82	255.36	2307.23

* All subjects read the material at least once.

to read," "which half more pleasant to read"). It was found that, despite instructions to the contrary, subjects' answers were based on content as well as style. This was shown by the fact that the 'cooling system' material (second half of the lesson) was judged easier and more pleasant to read than the 'induction system' material (first half of the lesson) whatever the style combinations used. A critical ratio of 8.77 was found, the probability of chance occurrence being less than .001.

Possible reasons for this include: (a) greater general familiarity with the concept of the 'cooling system' than with the concept of the 'induction system'; (b) greater 'content interest' value of the cooling system; and (c) unpleasant connotations of the term 'induction' for these newly inducted airmen.

The split versions, however, permitted equating for content in order to study the effect of style. Two 6 x 2 contingency tables were set up, split versions x lesson halves. They yielded chi-squares of 21.19 and 29.93 (both significant beyond the .001 level), showing that style had an effect on judgments of "easier to read" and "more pleasant to read" despite the influence of content.

Individual comparisons of the styles were then made, using the formula for the significance of the difference of percentages. Table 2 presents the data relevant to these analyses. It shows that the Easy style was judged significantly easier and more pleasant than the Hard, but not quite significantly easier or more pleasant than the Present. The Present style was judged significantly easier and more pleasant than the Hard. Tetrachoric correlation coefficients between judgments of easier to read and more pleasant to read ranged from .86 to .97.

Immediate Retention Test Score.—Analysis was next made of the

TABLE 2.—PERCENTAGES OF SUBJECTS WHO INDICATED ONE HALF OF LESSON EASIER AND ONE HALF MORE PLEASANT, BASED ON COMPARISONS OF SIX 'SPLIT' STYLE DIFFICULTY VERSIONS

Version*	Percentage Indicating Each Half Easier		Percentage Indicating Each Half More Pleasant	
	1st half	2nd half	1st half	2nd half
EP	39	61	30	70
PE	28	72	22	78
	C. R. = 1.53, P = .06 (point)		C. R. = 1.21, P = .11 (point)	
EH	44	56	45	55
HE	22	78	15	85
	C. R. = 3.07, P < .001 (point)		C. R. = 4.25, P < .0001	
PH	51	49	46	54
HP	34	66	34	66
	C. R. = 2.27, P = .01 (point)		C. R. = 1.62, P = .05 (point)	

* The first letter in each version designation represents the treatment of the first half, and the second letter the second half.

scores on the fifty-item test. A 3 x 1 analysis of variance of scores on the three unsplit versions was computed, followed by individual t-tests of mean score differences. Table 3 presents these data. It shows (a) significant variance attributable to style difficulty, (b) the Easy style resulted in a significantly higher mean score than the Present or Hard styles, and (c) the Present style resulted in a higher (but not quite significantly higher) mean score than the Hard style.

It had been found previously that subjects' mechanical aptitude indices (stanine scores) correlated highly with test scores on the preliminary tests developed for use with the lesson material. Therefore, correlations were computed between mechanical aptitude indices and test scores in this study. The correlations were .71 for the Easy, .75 for the Present, and .74 for the Hard versions. Since there were slight differences between the mean mechanical aptitude indices of the three groups, comparable stratified samples of one hundred Easy, Present and Hard subjects were pulled for analysis. A 3 x 1 analysis of variance and individual t-tests yielded comparable results to those obtained in the previous analysis.

Two further 3 x 3 analyses of variance were possible, utilizing test scores on the split versions. The first analysis involved the use

TABLE 3.—IMMEDIATE RETENTION TEST SCORE DATA, ANALYSIS OF VARIANCE AND *t*-TESTS FOR THREE UNSPLIT STYLE DIFFICULTY VERSIONS

Style Difficulty Version	Number of Subjects	Mean*	SD	Range
Easy	110	27.19	8.68	11–45
Present	110	24.69	8.23	9–45
Hard	110	23.07	7.70	6–48

Source of Variation	Sum of Squares	df	Mean Square	F
Between versions	947.03	2	437.52	6.97, .01 > P > .001
Within versions	22,209.90	327	67.92	

Versions Compared	*t*
Easy—Present	2.25, P = .012 (point)**
Easy—Hard	3.71, P < .001 (point)
Present—Hard	1.46, P = .072 (point)

* Maximum possible score, 50; chance score, 12.5.

** These values are based on the use of the normal probability table for greater exactness.

of scores on the first half of the test (which was based on the first half of the lesson material); the second analysis involved scores on the second half of the test. Table 4 presents the results of these analyses.

Table 4 shows that the variance attributable to style difficulty was again significant in both analyses. It also shows that individual mean score differences were comparable to those previously found.

It is possible in this sort of testing that items might be accidentally 'biased' or 'slanted' in favor of a particular version. If such slanting favored the Easy version, for example, mean score differences might be found due to large percentage passing differences on only a few of the total number of items. It will be recalled that using the Present version as the base for the test was an attempt to turn any bias which might exist in the conservative direction. Nevertheless, the subjects who had the Easy, Present and Hard versions were compared as to percentages passing each item as an additional check on bias.

TABLE 4.—ANALYSES OF VARIANCE FOR SIX SPLIT AND THREE UNSPLIT STYLE DIFFICULTY VERSIONS (989 SUBJECTS)*

Source of Variation	Sum of Squares	df	Mean Square	F
First-half Scores				
Style difficulty	217.82	2	108.91	6.56, .01 > P > .001
Position	7.76	2	3.88	.23, P > .05
Interaction	130.53	4	32.63	1.97, P > .05
Within groups	16,276.15	981	16.59	
Second-half Scores				
Style difficulty	540.23	2	270.11	10.63, P < .001
Position	75.53	2	37.76	1.49, P > .05
Interaction	208.33	4	52.08	2.05, P > .05
Within groups	24,916.13	981	25.40	

* Version PE had only one hundred nine subjects; in order to keep the N's equal in all versions and thus simplify analysis, the missing subject was assumed to have the mean score of the other subjects.

TABLE 5.—SIGN TESTS OF DIFFERENCES IN NUMBER OF ITEMS IN WHICH SUBJECTS HAVING ONE STYLE DIFFICULTY VERSION EXCELLED THOSE HAVING ANOTHER VERSION

Versions Compared	Number of Items			Critical Ratio
	Easy *excelled*	No *diff.*	Present *excelled*	
Easy vs. Present	37	2	11	2.82, P < .01 (level)
	Easy *excelled*	No *diff.*	Hard *excelled*	
Easy vs. Hard	44	1	5	5.51, P < .00001 (level)
	Present *excelled*	No *diff.*	Hard *excelled*	
Present vs. Hard	38	2	10	3.95, P < .001 (level)

Table 5 presents the data for number of items in which each group excelled. It can be seen that the effects of differences in style difficulty were spread throughout the tests, indicating little evidence of bias.

DISCUSSION

In discussing the results of this study, it might be well to mention briefly certain parts of the study which could not be reported for lack of space. The first involves preliminary work, using the same experimental materials, carried on in the classes of airmen in advanced technical training. It was found that background information on the lesson material, gained in previous training, was very high, and that experimental test score differences between style versions did not appear. Thus, while style difficulty appears to affect immediate retention of subjects who are naïve regarding material, subjects who have considerable knowledge of the material may profit little if any from an easier style of writing.

The second part of the study not mentioned previously in this paper is that regarding delayed retention. Subjects were retested (using the same fifty-item test) two weeks after original testing. Experimental differences, though still in the same direction (i.e., favoring Easy over Present and Present over Hard), had decreased below the five per cent level of significance. It is not known whether this was a function of style difficulty, of the level of difficulty of the test, of the delay between first and second testings (the two-week delay was necessary for administrative reasons; it was longer than desirable), or of other possible variables. An analysis of retention scores did suggest one other possible reason, however. In analyzing second-half test scores, it was found that the style treatment of the first half of the lesson material significantly affected second-half scores. Inspection of the means showed that a Hard first-half treatment resulted in higher mean second-half scores (for Easy, Present and Hard groups combined) than a Present first-half treatment, and Present than Easy. This suggests that beginning with a harder (and less pleasant) style may cause subjects to 'compensate', thus expending more effort on the material that follows. This same tendency, though not significant, was found in the analysis of split-half immediate retention scores (cf. position variance in the two analyses presented in Table 4).

This 'compensation' may be somewhat similar to that found by Laird (*8*) and Freeman (*4*), working with typists. It does seem to run counter to the general principle of 'breaking a subject in' on easy material. However, it should be noted that motivation was very high during the experimental program; lacking high motiva-

UNIVERSITY OF MICHIGAN LIBRARIES

tion, subjects might not show this behavior. Further work is required, especially since this result suggests that compensation may decrease experimental differences (i.e., between Hard and Easy styles). Certain tentative findings suggest this may have occurred in this study.

CONCLUSIONS

The results of this study indicate that an easier style of writing may result in (a) greater and more complete immediate retention, (b) a greater amount read in a given time, and (c) more 'acceptable' reading material. This was the case for the lesson used in this study, even though content, technical terms, and number of words were held constant.

The study also indicates, however, the great importance of content in determining how well material will be accepted. It shows, in addition, the high relationship between judgments of material as easier to read and more pleasant to read.

REFERENCES

1) E. Dale and Jeanne S. Chall. "A formula for predicting readability: Instructions." *Educ. Res. Bull.* (Ohio State Univer.), 1948, 27, 37–54.

2) R. F. Flesch. "A new readability yardstick." *J. appl. Psychol.*, 1948, 32, 221–233.

3) R. F. Flesch. "Measuring the level of abstraction." *J. appl. Psychol.*, 1950, 34, 384–390.

4) G. L. Freeman. "Changes in tension-pattern and total energy expenditure during adaptation to 'distracting' stimuli." *Amer. J. Psychol.*, 1939, 52, 354–360.

5) S. N. Hotchkiss and D. G. Paterson. "Flesch readability reading list." *Personnel Psychol.*, 1950, 3, 327–344.

6) G. R. Klare and B. R. Buck. *Know Your Reader*. New York: Hermitage House, 1954.

7) G. R. Klare, J. E. Mabry and L. M. Gustafson. "The relationship of verbal communication variables to immediate and delayed retention and to acceptability of technical training materials." Lackland AFB, Tex.: Air Force Pers. & Tng. Res. Center, Dec. 1954. *Research Bulletin.* (AFPTRC-TR-54-103.)

8) D. A. Laird. "Experiments on the physiological costs of noise." *J. Nat. Inst. Industr. Psychol.*, 1929, 4, 251–258.

9) C. E. Swanson and H. G. Fox. "Validity of readability formulas." *J. appl. Psychol.*, 1953, 37, 114–118.

PREDICTION OF THE ACADEMIC SUCCESS OF COLLEGE FRESHMEN

ROBERT A. JACKSON

Michigan State College

All new undergraduates at Michigan State College are required to take four tests: (1) ACE Psychological Examination developed by the Educational Testing Service, (2) Michigan State College Reading Test constructed by the Board of Examiners, (3) Test of English Usage, Third Edition, constructed by Benjamin B. Hickok of the Department of Communication Skills in coöperation with the Board of Examiners, and (4) Arithmetic Proficiency Test constructed by the staff of the Arithmetic Improvement Service. This study deals with the relationship between performance on these tests and academic success as measured by first-term grade point average. The individuals in this study were first-term freshmen at Michigan State College in the fall term of 1952. Only those carrying a full college load (from twelve to nineteen credit hours) were included.

Of the 3053 freshmen tested during Orientation Week, 2983 completed one term of college work; the other seventy students withdrew from college during their first term and received no college grades. Since it was known that the mean grade point average of the women was significantly higher than that of the men, the students were divided on the basis of sex. The means, standard deviations and intercorrelations of the five variables were obtained for the 1687 men and the 1296 women completing one term of college work; they are given in Tables I and II. The grade point average is calculated in the usual manner, with A assigned the value 4 down to F with the value 0; the scores on the four tests are derived scores based upon a ten-point scale calculated from a normal distribution by subdividing the ordinates of the normal curve from $+3\sigma$ to -3σ into ten equal units.

Tests of statistical significance were applied to the data given in Tables I and II and the following conclusions were apparent:

1) MSC freshman men have a significantly higher mean on the Arithmetic Test.

2) MSC freshman women have a significantly higher mean on the English Test and the Reading Test.

TABLE I—MEANS AND STANDARD DEVIATIONS ON THE INDICATED VARIABLES FOR MEN AND WOMEN

Variable	Mean		Standard Deviation	
	Men	Women	Men	Women
Grade Point Average	2.2	2.4	.70	.68
Psychological Examination	5.4	5.3	1.71	1.65
Reading Examination	5.2	5.5	1.70	1.64
English Examination	4.9	6.1	1.64	1.56
Arithmetic Examination	5.6	5.0	1.68	1.61

TABLE II—INTERCORRELATIONS BETWEEN THE INDICATED VARIABLES

(The values above the diagonal are for men, those below are for women)

	Grade Point Average	Psych. Exam.	Reading Exam. (Men)	English Exam.	Arithmetic Exam.
Grade Point Average		.43	.50	.49	.46
Psychological Examination	.52		.74	.56	.47
Reading Examination	.64	.74		.59	.39
English Examination (Women)	.52	.60	.61		.44
Arithmetic Examination	.49	.49	.42	.46	

3) MSC freshman women have a significantly higher mean grade point average.

4) The means of the men and women on the Psychological Test are not significantly different.

5) The best predictor of grade point average is the Reading Test.

Utilization of the multiple regression approach with two or more variables yielded no significantly better prediction of college success than did the Reading Test alone. The preceding data and other knowledge about the four tests indicated the desirability of restricting further investigation to the relationship of the Psychological Test and Reading Test to the grade point average.

The next part of this investigation was concerned with the relationship between grade point average and number of credits carried. The individuals in each of the sex groups were further classified into eight subgroups according to the number of credits

carried. The analysis of variance technique was used to test differences in group means. The data are not presented in this report but they indicated that the mean grade point average increased regularly with an increased credit load. The mean score on both the Psychological Test and the Reading Test also increased with credit load. The analysis of covariance technique was next used to adjust grade point averages on the basis of Reading Test scores. The adjusted mean grade point averages were still significantly different but at a much lower level of significance than the unadjusted grade point average means. This indicated one of the major causes of differences in mean grade point average to be differences in reading ability; however, other factors not included in this study are operating in such a manner as to cause the adjusted mean grade point average to increase with credit load. In other words, students taking heavier credit loads are generally more able students and continue to perform at the expected higher level despite the heavier load.

The mean scores on the Psychological and Reading Tests for those men completing one term and for those withdrawing during the first term are given in Table III; similar data for the women are given in Table IV. From Table III it is apparent that the men who withdrew during the first quarter had lower mean scores on both of the tests than the group completing one term. Table IV

TABLE III—MEAN SCORE ON THE PSYCHOLOGICAL AND READING TESTS FOR THE INDICATED GROUPS OF MEN

	N	Psychological Exam.	Reading Exam.
Men withdrawing during first term	48	4.56	4.38
Men completing first term	1687	5.38	5.21

TABLE IV—MEAN SCORE ON THE PSYCHOLOGICAL AND READING TESTS FOR THE INDICATED GROUPS OF WOMEN

	N	Psychological Exam.	Reading Exam.
Women withdrawing during first term	22	5.14	5.59
Women completing first term	1296	5.33	5.54

TABLE V—PERCENTAGE OF EACH ABILITY GROUP IN INDICATED GRADE POINT GROUP (MEN)

Ability on Psychological Examination	Grade Point		Ability on Reading Examination	Grade Point	
	Passing	Failing		Passing	Failing
High	94.6	5.4	High	96.6	3.4
Middle	78.7	21.3	Middle	80.2	19.8
Low	51.6	48.4	Low	48.8	51.2

TABLE VI—PERCENTAGE OF EACH ABILITY GROUP IN INDICATED GRADE POINT GROUP (WOMEN)

Ability on Psychological Examination	Grade Point		Ability on Reading Examination	Grade Point	
	Passing	Failing		Passing	Failing
High	97.5	2.5	High	98.6	1.4
Middle	86.6	13.4	Middle	87.0	13.0
Low	55.9	44.1	Low	48.7	51.3

shows that the women withdrawing from school during the first term had mean Reading and Psychological Test scores nearly identical with the mean scores of those completing one term. Apparently factors other than ability are operative in the case of women withdrawing from college. Of the freshmen tested during Orientation Week, 2.8 per cent of the men and 1.7 per cent of the women withdrew during their first quarter.

Correlations between test scores and grade point averages have already been given, but these are rather difficult to use directly. A second method of analyzing the relationship between grade point average and test scores involves the computation of the probability that persons with a given score on the predictor will achieve various levels of success on the criterion. This method helps to communicate to individuals the probability of doing at least as well as certain cut off points on the criterion. Since the critical grade point is a C average, it seemed desirable to determine the chances of various students doing at least average work in their first term of college. It was felt that many extreme situations in the first quarter of college might tend to result in the first-term grade point average being unrepresentative of the individual's eventual grade point average; therefore, the grade point distribution was dichotomized

at a grade point average of 1.75. Those individuals with a grade point average of 1.75 or above were called the passing group and those with a grade point average less than 1.75, the failing group.

To obtain percentages for the extreme groups on the basis of larger numbers, the ten-point scale was condensed into a three-point scale: (1) high-ability group (derived scores 10, 9, 8), (2) middle-ability group (derived scores 7, 6, 5, 4), and (3) low-ability group (derived scores 3, 2, 1). Two tables were set up with the dichotomized grade point forming one classification and these three ability groups the other. The entries in the tables are the per cent of each ability group that were in the passing or the failing group. The data for the men are given in Table V and for the women in Table VI.

Tables V and VI indicate that a much greater percentage of the low-ability group will obtain less than a C average than either of the other two groups. Practically all of the high-ability group do obtain an approximate C average or better. About fifty per cent of those individuals with a derived score of 3 or less will fail to make a C average the first term in college.

In interpreting these statistical data, the following limitations should be kept in mind:

1) A certain error of measurement pertains to each derived score; thus, the true score for a particular individual on one of the tests might be 6 rather than his observed score of 5, or 4 rather than 5.

2) The grade point average is based on one quarter of work; thus a particular student's grade point average may not be representative of the total grade point average he will eventually compile, due to possible extreme circumstances in this, his first quarter in college. For example, a student with high ability may do relatively ineffective work in certain areas due to lack of interest, motivation, etc.

3) The probabilities presented are based on group performance, and do not necessarily hold for individual prediction.

After considering these limitations, however, it seems that several statements are appropriate:

1) Women obtain significantly higher grades than men.

2) The Reading Test is the best predictor of academic success.

3) Women tend to perform more nearly in accord with their measured ability than do the men.

UNIVERSITY OF MICHIGAN LIBRARIES

4) In general, the percentage of an ability group failing to obtain at least a C average increases as the ability of the group decreases, but some students in the low-ability group make satisfactory grade point averages.

5) Practically all of the individuals in the high-ability group obtain at least a C average; whereas approximately fifty per cent of those in the low-ability group fail to obtain a C average.

A STUDY OF THE EFFECTS OF TEACHER KNOWLEDGE OF PUPIL CHARACTERISTICS ON PUPIL ACHIEVEMENT AND ATTITUDES TOWARDS CLASSWORK[1]

KENNETH B. HOYT

College of Education, State University of Iowa

To say that individual differences exist in school children has little significance simply as an observation. The significance lies in the actions one can take as a result of having made this observation and the results which ensue in terms of changes in students.

Many people prominent in education have, over the last fifty years, urged teachers to discover the ways in which and the degrees to which school children differ in various characteristics (*1*, *2*). Some have been so bold as to state specifically such beliefs as "There is no doubt that better teaching results from supplying teachers with pertinent facts about all their pupils." (*6*).

THE PROBLEM

One approach to the problem of the relative advisability or inadvisability of providing teachers with information concerning pupil characteristics is to study some of the effects resulting from such knowledge on the part of teachers. This investigation was concerned with studying the effects of teacher knowledge of pupil characteristics on pupil achievement and teacher-pupil relationships. Seven major questions were devised which served as the focus for the investigation. These were as follows:

1) What is the effect of providing teachers with systematic appraisal data relative to pupil characteristics on teacher knowledge of pupil characteristics?

2) What is the effect of teacher knowledge of pupil characteristics on pupil achievement?

---

[1] Summary of Ph.D dissertation presented to the Graduate Faculty at the University of Minnesota. The author is pleased to acknowledge the advice and encouragement which Professor Willis E. Dugan gave to this study.

3) Does the effect of teacher knowledge on pupil achievement act differentially for boys and girls?

4) Does the effect of teacher knowledge on pupil achievement act differentially on students from different learning ability levels?

5) What is the effect of teacher knowledge of pupil characteristics on pupil attitudes towards teachers?

6) Does the effect of teacher knowledge on pupil attitudes towards teachers act differentially for boys and girls?

7) Does the effect of teacher knowledge on pupil attitudes towards teachers act differentially on students from different learning ability levels?

The major experimental factors studied with reference to the treatment effect were teacher knowledge, sex, and ability level. The major experimental variables were pupil achievement and pupil attitudes toward teachers.

POPULATIONS STUDIED AND SAMPLING METHODS EMPLOYED

The experiment was conducted during the school year 1953–54 utilizing eighth-grade students in English, mathematics, and social studies classes from two public junior high schools in Minnesota. In School A, two hundred and ninety students were included on the roster and in School B, three hundred and fifty-six students. The actual experimental period was six months in length and all findings should be interpreted with this limitation in mind.

Individual students in each school were selected for and assigned to experimental sections by the method of random sampling within sex and scholastic aptitude sub-populations. Six sub-populations were formed from the roster in each school with boys and girls each being divided into highest, middle, and lowest ability groupings on the basis of obtained IQ scores on the *Otis Quick Scoring Mental Ability Test*, Beta level, Form Cm, which had been administered to all students in both schools during the school year 1952-53.

In School A, ninety students were selected and assigned at random, thirty to each of three experimental sections. Complete analysis data were available for eighty-five of these ninety students at the conclusion of the experimental period. In School B, one hundred and twenty students were selected and assigned at random, forty to each of three experimental sections, with complete anal-

TABLE 1—NUMBERS OF STUDENTS, BY SECTIONS, IN EACH STRATUM FOR WHICH COMPLETE ANALYSIS DATA WERE AVAILABLE IN SCHOOL A

	Section 8A		Section 8B		Section 8C		Totals	
	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys
High IQ	5	5	5	4	5	5	15	14
Middle IQ	4	5	5	4	6	4	15	13
Low IQ	5	5	5	4	5	4	15	13
Totals	14	15	15	12	16	13	45	40

TABLE 2—NUMBERS OF STUDENTS, BY SECTIONS, IN EACH STRATUM FOR WHICH COMPLETE ANALYSIS DATA WERE AVAILABLE IN SCHOOL B

	Section 8D		Section 8E		Section 8F		Totals	
	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys
High IQ	7	5	7	5	7	6	21	16
Middle IQ	5	8	6	6	7	6	18	20
Low IQ	6	6	6	6	5	5	17	17
Totals	18	19	19	17	19	17	56	53

ysis data being available for one hundred and nine students. The effects of the loss of these students as well as a graphical representation of the sub-populations involved can be obtained from an examination of Tables 1 and 2.

SPECIFICATION AND ASSIGNMENT OF THE TREATMENT VARIABLES

Three distinct treatments were applied to each section during the experimental period. These were arranged for both schools in such a way that any one teacher was responsible for a different treatment with each section. Definitions of the three treatment variables are presented in terms of the amount of information concerning pupil characteristics given to teachers.

The treatment of 'No Information' (N) is defined as the process of giving the teacher only the names of the pupils and no other systematic information.

The treatment of 'Test Scores' (T) is defined as the process of giving the teacher the pupils' test scores. These consisted of converted IQ scores from the *Otis Quick Scoring Mental Ability Test*, Beta level, Form Cm. (which all students took during the school year 1952-53); scaled scores, by parts, for the *Coöperative English Test*, Lower Level, Form Z; the *Coöperative Social Studies Test for Grades 7, 8, and 9*, Form X; and the *Coöperative Mathematics Test for Grades 7, 8, and 9*, Form X (which each student took in September, 1953.) No other systematic information was given the teacher.

The treatment of 'Tests Plus Other Information' (TO) is defined as the process of giving the teacher, for each student, all the information described for the Test Scores treatment plus a pupil information blank which was administered in September, 1953, and constructed by the experimenter. In addition, each teacher was encouraged to discover all he possibly could about the characteristics of each student receiving this treatment.

These treatments were assigned to experimental sections by subject using a table of random numbers compiled by Fisher and Yates (*3*). The resulting assignments for School A are shown in Table 3 and for School B in Table 4.

TABLE 3—ASSIGNMENT OF TREATMENT VARIABLES TO EXPERIMENTAL SECTIONS IN SCHOOL A

Section	Teacher 1 (English)	Teacher 2 (Mathematics)	Teacher 3 (Social Studies)
8A	TO	N	T
8B	T	TO	N
8C	N	T	TO

TABLE 4—ASSIGNMENT OF TREATMENT VARIABLES TO EXPERIMENTAL SECTIONS IN SCHOOL B

Section	Teacher 4 (English)	Teacher 5 (Mathematics)	Teacher 6 (Social Studies)
8D	T	TO	N
8E	TO	N	T
8F	N	T	TO

DESCRIPTION OF TESTS AND FORMS USED

Student achievement was measured in September and again in April using the *Coöperative English Test*, Lower Level, Form Z, the *Coöperative Mathematics Test for Grades 7, 8, and 9*, Form X, and the *Coöperative Social Studies Test for Grades 7, 8, and 9*, Form X. While these tests measure recall and recognition, it is apparent that what some would regard as the relatively more important goals of retention and transferability of learned materials to new situations is not measured by these instruments. This serves to specify one limitation of the study.

Information relative to other pupil characteristics which were considered as important aspects of treatment TO were collected in September by administration of the pupil information blank.

Student attitudes towards particular teachers were measured by use of an inventory constructed by the experimenter. It consisted of items taken from the *Student Reaction Inventory* (4), contained in category 7, "Individualization of Instruction," which, on an item analysis, showed differences between 'good' and 'poor' teachers (5) significant at the 1% level. Forty such items were selected and five more non-discriminating items which had high face validity were added making forty-five items in all. Responses typical of those students make about 'good' teachers were assigned a weight $+1$; those associated with 'poor' teachers a weight of -1; and all items answered 'undecided' were scored as zero. This instrument was administered to each student three times, once in each of the three experimental groups to which he was assigned.

Each teacher was asked to complete a blank consisting of questions designed to measure the degree to which he was able to recall specific characteristics of certain randomly selected pupils in the various experimental sections. Six students, one from each stratum, were picked at random from each experimental section in a given school and the three teachers involved were asked to complete the blanks for these eighteen students.

EXPERIMENTAL PROCEDURES

The experimental procedures employed in this investigation can perhaps be best described in terms of controls applied during the experimental period to the various treatment procedures. Rigid controls were also applied to the collection of data, but these are not described here.

UNIVERSITY OF MICHIGAN LIBRARIES

To maintain the limitations prescribed by treatment N teachers were urged, for each member of the section where this treatment was applied, to: (a) Refrain from examination of the pupil's cumulative record folder; (b) Refrain from discussion of this pupil in conversations with other school personnel; (c) Refrain from discussing with the student any topics which might lead to information concerning his activities outside of the classroom; and (d) Refrain from seeking out parents of the student for consultation and discussion.

To carry through the procedures prescribed by treatment T teachers were urged, for each member of the section where this treatment was applied, to maintain the same limitations described for treatment N. They were given copies of all initial achievement test scores and IQ scores for individual students in this section. In addition, graphs of distributions of these scores were supplied as was a scattergram showing the relationships between intelligence and achievement test scores in the subject the teacher was teaching.

To carry through the procedures prescribed by treatment TO the teachers were supplied, for each member of the section where this treatment was applied, with all of the information described for the T treatment. In addition, they were given: (a) Individual record folders for each student in which they were encouraged to place records of all information learned regarding characteristics of the student; (b) The pupil information blanks which the student had completed and on which the experimenter had marked some interpretive suggestions; and (c) A four-page group analysis of the section made on the basis of information obtained from the pupil information blanks.

To work towards a clear differentiation of treatments individual conferences were held with teachers at least once every other week during the experimental period. Here, in addition to providing frequent reminders of the restrictions imposed by Treatments N and T, the experimenter worked with each teacher in interpreting the significance of the data compiled for students in the T and TO sections. No attempt was made to force the teacher to discuss every pupil; rather, the choice of pupils to be discussed was always left with the teacher.

To work towards the prevention of gross changes in teaching method teachers were asked to: (a) Use the same textual material for all

experimental sections; (b) Cover the same units at approximately the same time with all experimental sections; (c) Use the same general approach in teaching method with all experimental sections, i. e., if a topic were presented by means of a student panel discussion in one section, this same approach should be used in all sections. Checks made during and at the conclusion of the experimental period satisfied the investigator that these restrictions were being carried out.

STATISTICAL DESIGN AND METHODS OF ANALYSIS

The statistical design utilized for analysis of achievement test scores is defined as $2 \times 2 \times 3 \times 3$ factorial. Specific factors involved included two schools, two sexes, three ability levels, and three treatments. This design was applied three times, once for each of the three school subjects—English, social studies, and mathematics. The method of statistical analysis employed was the analysis of variance and covariance with unequal frequencies in the subclasses using unweighted means. The method of unweighted means rather than the least squares method was chosen primarily because of the interest of the investigator in examining first-order interaction effects of schools, sex, and ability with the Treatment effects in the experimental design. Where analysis of covariance was employed, adjustment was made on the basis of initial achievement test scores.

The statistical design utilized for analysis of attitude test scores is defined as a $2 \times 3 \times 3 \times 3$ factorial. This design was applied twice, once for School A and once for School B. Specific factors included two sexes, three school subjects, three ability levels, and three treatments. The method of statistical analysis employed was the analysis of variance with unequal frequencies in the subclasses using unweighted means.

The statistical design utilized for analysis of teacher knowledge scores is defined as a 3×3 factorial with three teacher subjects and three treatments representing the factors involved. This design was applied twice, once for School A and once for School B. The method of statistical analysis employed was the analysis of variance with equal frequencies in the subclasses.

The assumptions involved in the use of these various statistical tools were tested separately for each design and found to be satisfied.

SUMMARY OF FINDINGS

1) Teacher knowledge of pupil characteristics was increased through the use of appraisal data. Significant F ratios were obtained on this analysis for mean treatment differences, for mean teacher-subject differences, and for mean interaction differences of treatment with the various teacher-subject combinations. The mean square for the treatment effect provided a significant F ratio when compared with the mean squares for error, teacher-subject, and interaction thus establishing itself as probably constituting a true effect. This finding indicates that the Treatment procedure did operate as intended.

2) Teacher knowledge of pupil characteristics was not found by itself to result in increases in pupil achievement in mathematics, social studies, or English. The application of the various treatment effects did not produce any significant differences in adjusted mean achievement test scores in mathematics and social studies. While significant adjusted mean differences in favor of Treatment N were found for adjusted achievement test scores in English, the complexities of the interaction effects present prevent one from regarding the observed differences as true treatment effects.

3) Teacher knowledge of pupil characteristics was not found to act preferentially on either sex in terms of its effect on increases in pupil achievement in mathematics, social studies, or English. There were no significant F ratios obtained for the interactions of the sex with the treatment factor.

4) Teacher knowledge of pupil characteristics was not found to act preferentially on ability levels of pupils in terms of its effect on increases in pupil achievement in mathematics, social studies, or English. The interaction effects of ability level with the treatment factor were not significant in mathematics and social studies classes. A significant F ratio was found for the interaction between ability level and treatment in English classes. However, because of significant interactions also present between treatment and schools and between ability and schools plus significant main effects for school and treatments, the exact interpretation of the significant interaction of ability with treatment was not clear.

5) There was a definite tendency for increases in teacher knowledge of pupil characteristics to improve pupil attitudes towards teachers. Significant differences in favor of Treatment TO were found in School B. In School A, the F ratio of 1.461 was smaller

than that of 3.04 required to reject the hypothesis of equality at the 5% level of significance. However, the mean of 8.37 for Treatment N, 8.77 for Treatment T, and 12.12 for Treatment TO were all in the direction indicated by the significant findings in favor of Treatment TO which were found in School B.

6) Teacher knowledge of pupil characteristics was not found to act preferentially on either sex in terms of its effect on pupil attitudes toward teachers. The interaction of sex with treatment was not significant in either school.

7) Teacher knowledge of pupil characteristics was not found to act preferentially on ability levels of pupils in terms of its effect on pupil attitudes towards teachers. The interaction of ability level with treatment was not significant in either school.

REFERENCES

1) Warren, Baller, *et al.* "Essential pupil-personnel records." *Teachers College Record*, January, 1947, 268–86.

2) R. N. Bush. "A study of student teacher relationships." *Journal of Educational Research*, May, 1942, 645–656.

3) R. A. Fisher and Frank Yates. *Statistical Tables for Biological, Agricultural and Medical Research*. New York: Hafner Publishing Company, Inc. 1949.

4) Paul R. Grim and Cyril J. Hoyt. "Appraisal of teaching competency" *Educational Research Bulletin*, April 16, 1952, 85–91.

5) ———. "Excerpts from two instruments for appraising teaching competency." *Journal of Educational Research*, May, 1953, 705–710.

6) Ralph D. Wadsworth. "Give teachers the facts about pupils." *California Journal of Secondary Education*, February, 1946, 84–87.

UNIVERSITY OF MICHIGAN LIBRARIES

BOOK REVIEWS

JOHN DOLLARD, FRANK AULD, JR., ALICE MARSDEN WHITE. *Steps in Psychotherapy*, New York, The Macmillan Company, 1953 pp. 222. $3.50

HARRINGTON V. INGHAM AND LEONORE R. LOVE. The *Process of Psychotherapy*, New York, McGraw-Hill Book Company, Inc., 1954, pp. 270. $5.00

O. HOBART MOWRER AND TWENTY-ONE CONTRIBUTORS. *Psychotherapy: Theory and Research*, New York, The Ronald Press Company, 1953, pp. 700.

CARL A. WHITAKER AND THOMAS P. MALONE. *The Roots of Psychotherapy*, New York, The Blakiston Company, 1953, pp. 236. $4.50

LEWIS R. WOLBERG. *The Technique of Psychotherapy*, New York, Grune and Stratton, 1954, pp. 869. $14.75

It might be appropriately asked why five books on psychotherapy should be reviewed in THE JOURNAL OF EDUCATIONAL PSYCHOLOGY. There are many reasons for this. In the first place, psychotherapy has been described over and over again by its protagonists as a process of learning. May, for instance, in Mowrer's book states: "The historical and logical analysis of the preceding pages indicates that psychological therapy is a learning, or more accurately, a relearning process." (p. 33) While Shoben in the same book states: "Psychotherapy may be fruitfully understood as a learning process." (p. 125) The cover of THE JOURNAL OF EDUCATIONAL PSYCHOLOGY specifically states that the journal is devoted primarily to the scientific study of problems of learning. Psychotherapy has also been defined as personality change and the JOURNAL is interested in the development of personality. The second reason for reviewing books on psychotherapy in this JOURNAL is that the principles of transference, countertransference and resistance apply equally to education as well as to psychotherapy. Children bring to the classroom situation attitudes toward the teacher and toward their classmates from their earlier experiences as incorporated in their personality trends which can be called transference attitudes. And teachers likewise bring to the classroom situation attitudes out of their own past which can be called countertransference. Children exhibit many kinds of resistance to learning and it is pos-

sible that the educator can profit from an understanding of some of the principles of the process of psychotherapy insofar as he can reinterpret these principles in terms of the educational process. In the third place, psychotherapy is coming more and more to be applied by college counselors, school psychologists, and even teachers to the problems of adjustment that their students present. In the fourth place, psychotherapists are beginning to find that the psychological tests used by the clinical psychologist are of some value to them—not only in order to assist them in making their preliminary diagnosis of the client's difficulties, but also, as Bellak has pointed out, as an integral part of the psychotherapeutic process itself. In the fifth place, many psychotherapists are now using methods that are essentially educational and the educator can perhaps profit from the experience of therapists with these methods.

The book by Dollard, Auld, and White might be called a supplement to the earlier book by Dollard and Miller, entitled *Personality and Psychotherapy*.[1] This new book contains the report of one case of Mrs. 'B', which serves to illustrate the principles and practice of psychotherapy as they were enunciated more formally and abstractly in the earlier book by Dollard and Miller. The process of psychotherapy which this case presents is a strange combination of nondirectiveness and directiveness. The patient is almost ordered to free associate, and the authors even recommend the use of mild threats in order to steer their client's responses along the lines of free association. In the third interview, we are told: "The therapist observed that with this hour the patient would start free association. Mrs. 'B', replied, 'I do not feel like talking'." Because of this refusal to begin free associating, Dollard in his supervisory comment suggests saying the following: " 'You have just a little time here. You have to decide how to make use of it. I think you are in sufficient trouble so that you should seriously consider keeping your appointments here'! You must motivate her to come and remind her of the misery that she is bearing." (p. 41)

Dollard's therapeutic process might be called an aggressive form of nondirectiveness. As supervisor, he instructs his therapists to proceed without hesitation in his tasks of labeling and interpreting. For instance, he suggests the following: "The ideal procedure in therapy would be first to establish the fact of the conflict, then to

[1] John Dollard and N. E. Miller. *Personality and Psychotherapy*. New York: McGraw-Hill Book Co., 1950.

recur to its origin. Being able to label the feeling and behavior appropriately could have a great effect." (pp. 109, 110) Apparently Dollard believes that the therapist should do this labeling and interpreting with assurance and that the therapist should expect to do most of the analytic work himself.

In reporting the case of Mrs. 'B' the upper part of each page is a report of the therapist who conducted the case, and the lower part of the page contains the suggestions made by the supervisor. This enables the reader not only to follow the treatment of the case, but also the supervisor's explanatory and interpretive comments with regard to the therapist's approach.

Part II of the book gives the results of psychological tests given to Mrs. 'B' and comments on their value. The authors come to the conclusion: "For our own part, we do not believe that the test reports would have been very helpful." (p. 197)

The book by Ingham and Love is an excellent example of teamwork between a psychiatrist and a clinical psychologist. The senior author is a psychiatrist attached to the student health service of the University of California at Los Angeles, and the junior author is a clinical psychologist on the staff of the Los Angeles Psychiatric service. No indication is given as to the relative contribution of these authors to the text. There is good reason to believe, however, that Miss Love conducted the second of the two cases presented in the appendix. This is an elementary and readable text on the process of psychotherapy. One gains the impression that the authors are insightful persons and that they are sensitive to the relationships between a therapist and his client. Technical vocabulary is avoided and the book is self-contained with no bibliography or footnote references to the work of other individuals.

The book by Whitaker and Malone on the other hand is much more of a scholarly treatise. The first author is probably the practitioner and is interested in process and techniques. The second author is more the scientist and is concerned with the relations between psychotherapy and biology. The book itself uses a somewhat unfamiliar vocabulary; and in spite of the fact that there is a glossary of terms at the end it is difficult reading because even familiar terms are used with a somewhat unfamiliar meaning. These authors are in the business of training psychotherapists at the Emory University School of Medicine, and the report is in part based on their own innovations. Whereas the book by Ingham and Love is ad-

dressed primarily to therapists dealing with the milder problems such as would be found in a college population, Whitaker and Malone are dealing with the problems of deeply regressed, probably schizophrenic individuals; and, consequently, psychotherapy as they envision it departs much more from ordinary counseling. One of the authors has made a distinct contribution in showing how personality as well as the physical body can be depended upon to have its own process of repair or readaptation, and that it is this process of seeking better adaptation and adjustment that every psychotherapist must depend upon as the inherent healing process.

One of the major contributions of this book has to do with new developments in the theory of countertransference. More than any other writer they show how the therapist brings his own problems into the psychotherapeutic situation. They introduce the interesting concept of the therapist becoming the patient on occasion when his own problems become intense and the patient reversing the rôle and serving as a therapist to the therapist. They also introduce the concept of the 'sliver problem'. These slivers represent areas of minor transference difficulties which have not been worked through in the therapist's own psychotherapy. "In the ordinary person, these are unimportant interpersonally since they do not involve critical areas. To the professional therapist, however, because of the multiplicity of his therapeutic relationships, they give rise to frequent and varied interpersonal *impasses*. During therapy these are continually being worked through by the therapist-vector in the patient, and they represent a major source of satisfaction to the therapist. Each time an *impasse* is resolved, such a sliver problem has probably been resolved in the process." (p. 151) These authors point out how the practice of psychotherapy contributes to the therapist's own readjustment.

The book edited by Mowrer is difficult to cover in the small space which can be allotted to it in this review. All of the contributors are clinical psychologists and all but two of them have the Ph.D. degree. One is impressed, however, by the great diversity of approaches and points of view made by these clinical psychologists; an outside observer must be impressed by the fact that psychotherapy means many different things to clinical psychologists. In the first chapter, May emphasizes the point that it is impossible for a therapist to avoid introducing his own personal values into the therapeutic situation. This point of view is contrary to that

held by Rogers in Chapter 2 who, as is well known, is an out-and-out nondirectivist.

Mowrer has taken opportunity to elaborate further his theories concerning neurosis in five of the chapters of the book of which he is the editor. This reviewer has searched diligently in order to find out exactly what psychotherapy means to Mowrer; and he finds it a little discouraging to read: "Psychotherapy is a complex procedure and there are many valid and suggestive ways of describing it. . . . No attempt will be made to describe this procedure comprehensively." (p. 148) One would expect something better than this in a volume of seven hundred pages. It is clear that Mowrer believes "that psychotherapy is not that of emotionally re-educating the patient, but of helping establish problem-solving habits which will enable emotions to operate as they are normally intended to." (p. 149) Drawing on his two-factor theory of learning, Mowrer states: "Although the neurotic's principal complaint is that of emotional suffering, the most effective therapeutic attack is made in the area of problem-solving behavior." (p. 148) One might conclude from this statement that Mowrer's psychotherapeutic approach is essentially directive, that is, in helping his patient forward in his problem-solving behavior. However, Mowrer contradicts this conclusion by stating: "Attempts to deal with the neurotic's problems either directly or directively have not brought very satisfactory results, and we can now see that in some ways these approaches probably perpetuate the strategies which constitute the very essence of neurosis." (p. 148) Elsewhere, Mowrer describes therapy as follows: "The person is invited to *talk*. . . . With an appropriate degree of tentativeness and timeliness, the therapist then begins to report *his* associations to the patient; and it is these which constitute the essence of 'interpretation' and which set in motion forces which are designed to arrest and reverse the dissociative trends within the patient." (p. 79) "The associations of the therapist serve to reinstate the conflicts which the repression resolved." (p. 79). "Instead of seeking to advance the patient's psychological welfare by dissolving a supposedly too severe superego, a therapist operating within the present frame of reference *accepts* the patient's high standards as evidence of good character and works toward the goal of making these standards more understandable and helpful, so that they may ultimately be harmoniously integrated with the rest of the personality. Consistent with this point of view, the

therapist also takes the attitude that depression, inferiority feeling, anxiety, and the other painful affects which the neurotic experiences are *signs of strength* within the personality." (p. 91)

In spite of the fact that Mowrer believes that the task of psychotherapy is not that of emotionally re-educating the patient, he has included a chapter (19) on physiological changes during psychotherapy by O'Kelly. This would indicate that Mowrer either is very broadminded and is willing to include all points of view, or he himself failed to recognize the inconsistency. Part II of this book is a collection of papers dealing with research methods and results that have bearing on psychotherapy. The research-minded psychologist will find a mine of information in these chapters dealing with research procedures currently being experimented with in testing some of the outcomes of psychotherapy.

The massive volume by Wolberg stands apart from these others. For the first time we have a comprehensive manual dealing with psychotherapy in its many aspects. This book undoubtedly will serve a as text in institutes for the training of psychotherapists and as a manual and reference book by established practitioners. Part I of the book deals with the general principles of psychotherapy. Wolberg divides therapy into three main types: supportive therapy, re-educative therapy, and reconstructive therapy. The first of these is essentially environmental manipulation; the second is what is commonly known as counseling; whereas the third is depth therapy which results in personality reorganization. Three chapters (5, 6, and 7), describing the variations under each of these headings constitutes a remarkably helpful survey of the different schools of psychotherapeutic thought, and the student will find Chapter 7 of value in orienting him to the different schools of psychoanalysis and neo-Freudianism. Parts II, III, and IV of the book deal with the beginning, middle and terminal phases of treatment. Of particular value are the copious case materials which illustrate adequately the theoretical points made in the text. In Part V special aspects are dealt with including the handling of emergencies, the problem of supervision, a chapter answering questions which therapists frequently ask, and the problems of recording. Chapter 50 dealing with psychotherapy in special conditions stands apart from the others. This reviewer would not find it difficult to believe that this chapter was written before the rest of the book for it represents a point of view about psychotherapy which is not consonant with

the main parts of the book since it represents an older and more strictly old-fashioned psychiatric approach. For instance, in the section on the treatment of alcoholism we read: "Sometimes a belittling and challenging manner mobilizes in the patient a need to prove that he has the 'guts' to master the craving for alcohol." Wolberg adds, however: "This, however, cannot be depended upon." (p. 616)

In spite of the fact that Wolberg is eclectic in this book and gives fair treatment to many points of view, his own approach to psychotherapy is essentially that of encouraging his patient to think through the causes for his difficulty. Certainly in the case illustrations Wolberg is more active than one would expect to find a psychoanalyst. For instance, one of his patients says: "Yes, I see what you mean doctor. If I can understand why I feel the way I do everything will be all right?" (p. 701) Then later Wolberg responds to this patient: "The only way you can change is to find out how the patterns you have been accepting as normal are abnormal." (p. 753) Apparently, therefore, Wolberg puts the entire emphasis in his therapeutic work on understanding and insight with the belief that when these are achieved the patient has overcome the greatest obstacles toward his recovery. In his activity he actually goes so far as to press his patients to make their own explanations for their difficulties. "Do you have any idea what caused these feelings, what is actually behind these symptoms?" (p. 701) It is not clear whether Wolberg actually expects his clients to give reasonable explanations for their difficulties or whether he is simply orienting them to search for and to be alert to the causal factors. His own comment to the remark just quoted is: "Focusing her attention on the fact that there are causes that underlie the symptoms." (p. 701)

There are passages in this book that should be of particular interest for educators. For instance, on page 128 and in the following three pages there is a discussion of the teacher in psychotherapy. This treatment is by no means dogmatic and while Wolberg believes that the teacher is not equipped to do psychotherapy in the formal sense, teachers can profit a great deal in understanding some of the principles of psychotherapy in dealing with emotional problems that their pupils present.

The chapter on supervision in the psychotherapeutic process has immediate application to the problem of supervision in education and the dynamic principles of supervision which Wolberg enun-

ciates applies to supervision in education as much as to supervision in psychotherapy. On page 650 Wolberg gives a list of fourteen propositions involved in the process of learning of psychotherapy, and the student of learning should gain from these valuable suggestions about some of the dynamics of the learning process. In passing, it may be added that Wolberg gives a very fair and well-informed discussion of the problems relating to the practice of psychotherapy by clinical psychologists. PERCIVAL M. SYMONDS

Teachers College, Columbia University

GEORGE R. KLARE AND BYRON BUCK. *Know Your Reader*. New York: Hermitage House, 1954, pp. 192. $2.95

Presumably, writing is done to communicate with a reader. Unless a writer knows the reader and his interests, he may well end up by talking to himself or to nobody. But if he knows his reader, he is able to write with some consideration for him. The authors of this treatise state that they have concentrated on kinds of information that can help a writer to meet his reader. Much of the material discussed is derived from application of readability formulas. Advantages as well as limitations of the formulas are noted. In the introductory material, consideration is given to the wide range of application of formulas along with some results.

The next series of chapters deal with the background for, and history of, readability research. Next comes a discussion of the reader himself. Without this, readability formulas are of little use. A series of three chapters deal with the technical background necessary for understanding formulas and their place in a writer's equipment. This is followed with discussions of preparation for writing and some aspects of style that make writing readable. Too little attention is given to format in relation to readability. Note is made of pitfalls which can arise from misunderstanding of formulas and their misuse. It is emphasized that even when formulas are employed properly, they leave unmeasured many of the important aspects of readable writing. A final chapter discusses the trend of contemporary reading research. Much of the material presented here is concerned with viewpoints and theory rather than research.

Certain statements in text as well as examination of the bibliography suggest that the authors have failed to note some of the recent experimental findings on scientific typography. Thus, (p.

132) the statement that all-capitals and italics tamper very little with set reading habits; (p. 134), that italics are no harder to read than ordinary Roman type; and (p. 133), that differences in perceptual span are relatively unimportant, as are differences in interfixation speed of movement.

The authors have presented a well-organized, well-written treatise. Their emphasis upon knowing the reader if communication through writing is to be effectively achieved is basic. Their evaluation of readability formulas and of the use and limitations of such formulas is a contribution of high value. Writers of textbooks, trade journals, materials for communications to workers in industry, and other books and magazine articles should profit by careful attention to the contents of this book.

MILES A. TINKER
University of Minnesota

C. C. Ross. *Measurement in Today's Schools.* Revised by Julian C. Stanley. (Third Edition). New York: Prentice-Hall, Inc., 1954, pp. 485.

The Ross text first appeared in 1941, and was revised in 1947. Following Professor Ross' death, Dr. Stanley prepared this third edition. No major changes in the pattern of the book or the point of view have been made since the first edition. The largest number of chapters bear on the use of test results, with two chapters on test construction, one on statistics, and one on characteristics of tests. In contrast to many other texts in this field, no space is given to systematic coverage of published tests or to problems associated with testing various content areas or types of behavior. Chapters on graph-making and public relations are offered.

Students have liked the Ross text for its definiteness. It gives specific procedures for item preparation, and covers many topics of interest to administrators and supervisors. On the other hand, the text seemed conservative even at its first appearance, and the revisions have preserved this character. The student is told many things which must seem to him purely academic in interest, and some chapters seem so to the reviewer also. The style is one of old-fashioned scholarship, with long and uninspired quotations, numerous tables, and footnotes by the dozen. On the positive side, Ross' treatment was sound, in the light of the knowledge of the 1930's.

Stanley has clearly undertaken to change the book as little as

possible. He has added new footnote references, has greatly modernized the chapter reading lists, and has added paragraphs at a few points. Several illustrative tables have been replaced with better examples. He has judiciously removed paragraphs by Ross, but has never rewritten material even where Ross' view is questionable. One change in the text increases an unfortunate overemphasis on item analysis as a way of determining test content. The interpretation of correlation is damaged by introducing a chart in which the Fisher *z* is taken as showing "the amount of relationship represented" by a given *r*, and by omitting Ross' reference to Taylor and Russell. To compensate for the absence of new ideas in the main chapters, Stanley has added one chapter on present trends. This material is too cursory to be significant; one simply cannot deal intelligently with factor analysis in five sentences, no matter how well chosen. This chapter does provide reassurance, lacking elsewhere in the book, that the testing field has not stagnated in recent years.

The chief fault in Ross' book is one which could not be remedied by retouching. The book is oriented around the superficies of measurement, such as scoring of rearrangement items and using legible record forms, rather than around behavior and learning. This is made obvious by the fact that the objectives of instruction and their relation to testing are discussed only in one three-page section. Curricular validity is dismissed with the trivial and incredible comment that this idea has "shifted the center of gravity from the curriculum to the child."

Those instructors who have been satisfied with the Ross text will be grateful for a revision such as might have been expected from Ross himself. The revision will not, however, appeal to any new audience.

LEE J. CRONBACH

University of Illinois

UNIVERSITY OF MICHIGAN LIBRARIES